THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 1 January 1991

Contents

(ISSN 0002-9890)

ARTICLES

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Shirley A. Hill JAMES T. FEY AND KENNETH M. HOFFMAN 1

Conway's Challenge Sequence COLIN L. MALLOWS 5

LETTERS TO THE EDITOR 21

UNSOLVED PROBLEMS

A Dynamic Puzzle MOSHE ROSENFELD 22

NOTES

A Continued Fraction and Permutations With Fixed Points HENRI DARMON AND JOHN McKAY 25

Smoothness from Finite Points VICTOR NEVES 27

Anthropomorphic Polygons GODFRIED TOUSSAINT 31

Inequalities for Areas Associated with Conics W. A. DAY 36

THE TEACHING OF MATHEMATICS

The Derivative á la Carathéodory STEPHEN KUHN 40

Positive Definite Matrices and Sylvester's Criterion GEORGE T. GILBERT 44

The Gibbs Phenomenon for Piecewise-Linear Approximation J. FOSTER AND F. B. RICHARDS 47

Teaching and Testing Mathematics Reading CARL C. COWEN 50

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 54

Advanced Problems and Solutions 63

REVIEWS

Lecturers on Minimal Surfaces. By Johannes C. C. Nitsche ROBERT OSSERMAN 72

TELEGRAPHIC REVIEWS 75

ACKNOWLEDGEMENTS 87

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia. PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $32 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second-class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

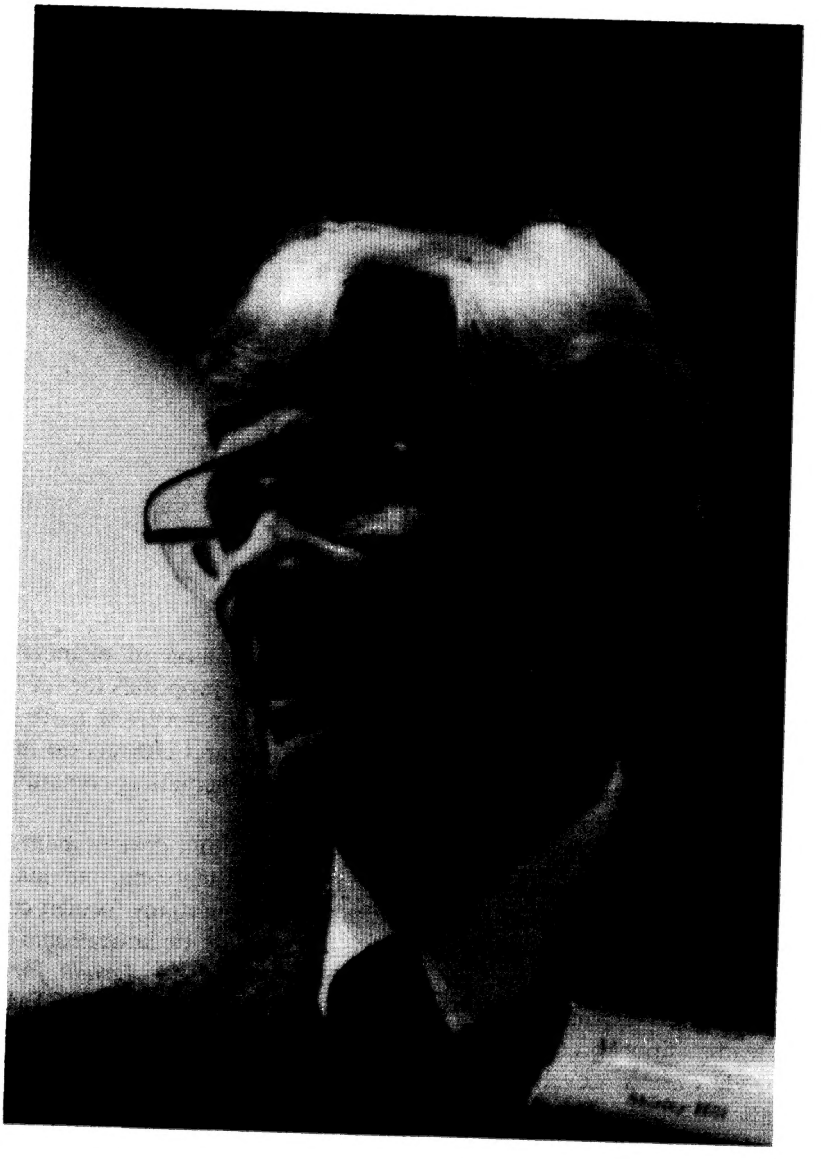
Award for Distinguished Service to Dr. Shirley A. Hill

JAMES T. FEY AND KENNETH M. HOFFMAN

During the past 30 years, no subject in the curriculum of elementary and secondary schools has received more critical attention or proposals for reform than mathematics. No subject in the college curriculum has received more focused attention. Throughout that crucial period of challenge and change, no individual has provided more distinguished leadership for the mathematical community than Shirley A. Hill.

Shirley Hill has played key roles in dozens of curriculum development projects and policy boards that have shaped the character of teaching and learning in mathematics at all levels, across the United States and in many foreign countries. It is of more than symbolic significance that the decade of educational ferment through which our country has just passed began with *An Agenda for Action*, developed by the National Council of Teachers of Mathematics (NCTM) under Shirley's leadership as president, and ended with the even more ambitious agenda of *Everybody Counts*, developed during her tenure as the first chairman of the Mathematical Sciences Education Board (MSEB).

Events have given our discipline the leading role in implementing new national education goals during the 1990s. Surely no single individual can be credited with having placed us in this challenging position. Yet, we can say with equal certainty that as a discipline we could not have come this far without the remarkable leadership of Shirley A. Hill. It is an especially appropriate time to recognize her contributions through the Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service.



Shirley A. Hill

A Phi Beta Kappa graduate of the University of Missouri, Shirley Hill taught in the schools of Kansas City, Missouri, before heading west to earn a Ph.D. from Stanford University in 1961. While at Stanford she became the assistant director of the Stanford University Arithmetic Project. This was the first in a series of contributions to prominent curriculum development projects that included the School of Mathematics Study Group, the African Mathematics Project, a project for teaching mathematics by radio in Nicaragua, and the Comprehensive School Mathematics Project.

In 1967, Shirley joined the Teacher Training Panel of the Committee on the Undergraduate Program of the Mathematical Association of America (MAA). She helped draft and promote the adoption of CUPM guidelines for the education of mathematics teachers—a problem area in which she has maintained a strong interest throughout her career.

Teacher education has been the focal point of her work as a professor, since 1963, at the University of Missouri-Kansas City. Her skill and dedication to this work have been widely recognized, including awards as Missouri Mathematics Teacher of the Year in 1978 and Amoco Award for Excellence in Undergraduate Teaching in 1981.

The recognition of Shirley's talents grew quickly, and in 1972 she was elected to the Board of Directors of the National Council of Teachers of Mathematics and the Board of Governors of the MAA. She chaired the National Advisory Committee on Mathematical Education and was a principal author of its influential report *Overview and Analysis of School Mathematics K-12*. In 1976 she was appointed chair of the United States Commission on Mathematical Instruction and the U.S. representative to the International Commission on Mathematical Instruction. In 1978 she was elected to a two-year term as president of the National Council of Teachers of Mathematics.

During the 1980s, Shirley served as a consultant to more than 15 states and to a broad array of curriculum development, assessment, and teacher education projects, including the University of Chicago School Mathematics Project, the National Assessment and Second International Study of Mathematics Achievement, and the National Board for Professional Teaching Standards. She was instrumental in the creation of the Mathematical Sciences Education Board and as chairman led its development from 1985 through 1989.

While serving in so many active and advisory roles on diverse projects and boards, Shirley has also managed to publish widely. Her contributions have ranged from *First Course in Mathematical Logic* (with Patrick Suppes) to the text, *Elementary Geometry*, for teachers (with Vincent Haag and Clarence Hardgrove) and chapters in the yearbooks of the National Council of Teachers of Mathematics, the NSSE, and the National Education Association. She was an associate editor of *The American Mathematical Monthly* from 1973–1977, and a member of the editorial panel for the 1973 NCTM Yearbook, *The Slow Learner in Mathematics*.

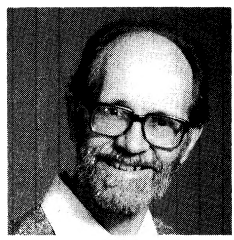
The numerous, elected, and appointed positions on national and international professional policy and advisory boards, and her frequency selection as leader of those groups, are unmistakable evidence of the respect that Shirley has earned throughout the mathematical community. She has, for many years, been a highly sought speaker at local, regional, national, and international professional meetings, where she conveys an inspirational sense of dedication to mathematics teaching.

In 1987, her institution, the University of Missouri at Kansas City, recognized her accomplishments by appointing her to the distinguished Curator's Professorship, through which she carries on both the dedicated teaching and the outstanding leadership for which we honor her.

Conway's Challenge Sequence

COLIN L. MALLOWS, *AT&T Bell Laboratories, Murray Hill, New Jersey 07974*

COLIN MALLOWS has been doing research in statistics at AT & T Bell Labs since 1960. He sees his main challenge to be that of making his subject seem less like sorcery.



1. Introduction. In a lecture at AT&T Bell Labs [1], J. H. Conway introduced the sequence

$$1, 1, 2, 2, 3, 4, 4, 4, 5, 6, 7, 7, 8, 8, 8, 8, 9, \ldots$$

defined by the recursion (for $n \geqslant 3$)

$$a(n) = a(a(n-1)) + a(n - a(n-1)).$$

He had proved that $a(n)/n \to 1/2$, but admitted to being unable to establish the rate of convergence. He offered a modest prize for determining the rate, and a most immodest one for the exact n at which $|a(n)/n - 1/2|$ last exceeds $1/20$. Notice that this is not a "find the next term" problem. Here we already know the rule for generating the sequence; the challenge is to develop enough understanding of its structure so that rigorous proofs can be given. To claim the larger prize it would not be enough merely to find a number that "looks like" the right answer. No (finite) amount of computation will suffice to prove that apparent regularities persist indefinitely.

It turns out, as we shall see, that this sequence does have much hidden structure. In fact, there are simple rules that work much faster than one-step-at-a-time; there is even a formula that enables us to compute $a(n)$ for any n. However it is not immediately clear how to prove these results.

What follows is an account of my search for structure in Conway's sequence, culminating in a proof that the structure that is found does in fact persist. Then there is an explanation of how Conway's number was calculated. Finally, there are comments on some related problems.

My own interest in this problem is two-fold. First, there is the simple pleasure to be found in taking up a challenge. But also, since I am by profession a statistician, I am interested in the process by which hidden structure (usually in statistical data, rather than a mathematical puzzle, as in this case) can be found. John Tukey and I have remarked [2] that success in an exploratory investigation is crucially dependent on

(i) a willingness to collect and study the data,
(ii) use of diagnostic techniques to show the unexpected,
(iii) an ability to recognize striking patterns,
(iv) enough understanding of the context of the problem to enable these patterns to be recognized as potentially meaningful,

(v) avoidance of precipitate commitment to models of clearly inadequate complexity, and
(vi) energetic following-up of the clues obtained.

All these components, except possibly (v), operate in the present case.

2. First steps. Let us look at some values of $a(n)$.

$n =$																					
1–20	1	1	2	2	3	4	4	4	5	6	7	7	8	8	8	8	9	10	11	12	
21–40	12	13	14	14	15	15	15	15	16	16	16	16	17	18	19	20	21	21	22	23	
41–60	24	24	25	26	26	27	27	27	28	29	29	30	30	30	31	31	31	31	32	32	
61–80	32	32	32	32	33	34	35	36	37	38	38	39	40	41	42	42	43	44	45	45	
81–100	46	47	47	48	48	48	49	50	51	51	52	53	53	54	54	54	55	56	56	57	

Some striking patterns are readily apparent. The differences

$$\Delta a(n) = a(n) - a(n-1)$$

are always 0 or 1, and this property is clearly persistent; if $\Delta a(n-1) = 0$ then $a(a(n-1)) = a(a(n-2))$ so the first component of $a(n)$ stays constant, while if $\Delta a(n-1) = 1$ then the second component of $a(n)$ stays constant. This shows that $a(n) < n$ for all $n \geqslant 3$, so that the sequence is well-defined.

Next, we notice repeated occurrences of 2's, 4's and 8's; in fact the last 2 occurs at $n = 4$, the last 4 at $n = 8$, etc. Applying strategy (ii) above, when we find anything, we should try to "subtract it out" so that a deeper level of structure can be seen. Here, we look at the function

$$A(n) = 2a(n) - n$$

which will be zero when n is a power of 2. Now successive values will differ by ± 1. Notice that

$$A(n) = A(a(n-1)) + A(n - a(n-1)).$$

FIGURE 1 shows the first 384 values of $A(n)$. It shows that $A(n)$ is positive and (surprise!) symmetric between its zeroes! If the positivity (and the humps!) persist, the values of $A(n)$ in any hump will be generated from the values in the immediately preceding hump; there will be no longer-range dependence. Let us number the humps so that the kth hump extends from 2^k to 2^{k+1}.

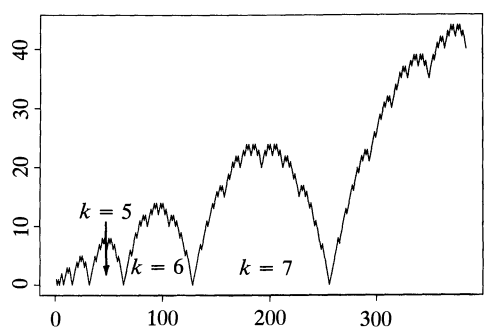


FIG. 1. $A(n) = 2a(n) - 1$ for $1 \leqslant n \leqslant 384$.

It is not clear how a hump is generated from the previous hump. The process is not local. For example, $a(100)$ (in the sixth hump) is computed as $a(a(99)) + a(100 - a(99)) = a(56) + a(44) = 31 + 26 = 57$, with the two components being far apart in the fifth hump.

Let us look at the heights of the humps in the middle.

$n = 3$	$A(n) = 1$	$n = 6$	$A(n) = 2$
12	2	24	4
48	6	96	12
192	20	384	40

We recognize the sequence $1, 2, 6, 20, \ldots$ as being central binomial coefficients, with $2, 4, 12, 40, \ldots$ being just double these. At this point it is worth computing more terms in the sequence, to see whether these regularities persist. FIGURE 2 shows 6144 terms of $A(n)$, and indeed they do, at least this far.

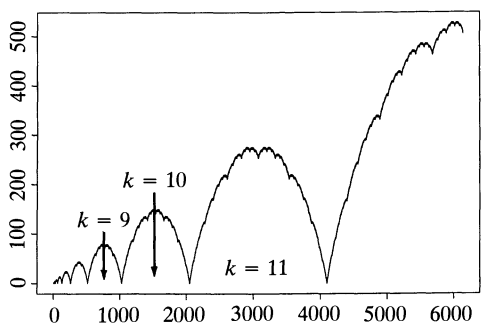


FIG. 2. $A(n)$ for $1 \leq n \leq 6144$.

3. The breakthrough. How can we describe the shapes of the humps? Let us define

$$D(n) = A(n) - A(n-1) = 2\,\Delta a(n) - 1$$

and

$$D_k(j) = D(2^k + j) \quad \text{for} \quad 1 \leqslant j \leqslant 2^k.$$

Some values are

D_3: 1 1 1 −1 1 −1 −1 −1
D_4: 1 1 1 1 −1 1 1 −1 1 −1 −1 1 −1 −1 −1 −1

etc., which are conveniently encoded as

d_3:	3 −1	1 −3							
d_4:	4 −1	2 −1	1 −2	1 −4					
d_5:	5 −1	3 −1	2 −1	1 −2	2 −1	1 −2	1 −3	1 −5	
d_6:	6 −1	4 −1	3 −1	2 −1	1 −2	3 −1	2 −1	1 −2	(negative reverse)

We notice that d_k has 2^{k-1} elements and is skew-symmetric. Also the sum of the absolute values of the elements is 2^k (we knew that!). However, it is not clear how one d can be derived from the previous one. (Even assuming that this is what

continues to happen!) Earlier, we found a hint that odd and even humps should be considered separately, so let us compare d_3 with d_5, and d_4 with d_6, etc. Now we find another regularity: for $k = 3, \ldots, 9$, d_{k+2} can be obtained from d_k by the following operation, which we call Rule L (for Local).

Rule L.
(i) replace each element x in d_k by the string of elements $l(x)$ where

$$\begin{array}{lllllll} l(1) = 2\ -1 & 1 & & & & & \\ l(2) = 3\ -1 & 2 & & & & & \\ l(3) = 4\ -1 & 3\ -1 & 2 & & & & \\ l(4) = 5\ -1 & 4\ -1 & 3\ -1 & 2 & & & \\ l(5) = 6\ -1 & 5\ -1 & 4\ -1 & 3\ -1 & 2 & & \\ L(6) = 7\ -1 & 6\ -1 & 5\ -1 & 4\ -1 & 3\ -1 & 2, & \end{array}$$

etc.

$$l(-x) = -\text{reverse of } l(x).$$

(ii) increase the first element of the new string by 1 and decrease the last element by 1.

Notice that $l(1)$ and $l(-1)$ are anomalous. Of course we have not yet proved that this rule works, we have only observed that it does in the first seven instances.

4. We need more insight. It is very awkward that Rule L connects the kth hump (from 2^k to 2^{k+1}) with the $k + 2$nd hump. The step to the $k + 1$st hump seems more complicated. Let us look at how D_{k+1} comes from D_k. Define (for $n \geqslant 3$)

$$F(n) = A(a(n-1)) - A(a(n-2))$$

$$G(n) = A(n - a(n-1)) - A(n - 1 - a(n-2))$$

so that

$$D(n) = F(n) + G(n)$$

and $F(n), G(n)$ are zero or ± 1. We see that the definition of Conway's sequence implies that for all $n \geqslant 3$,

$$F(n)G(n) = 0$$

while the successive nonzero values of $F(n)$ reproduce $D(n)$, and the successive nonzero values of $G(n)$ also reproduce $D(n)$. Let us write

$$F_k(j) = F(2^k + j)$$

$$G_k(j) = G(2^k + j).$$

To establish that the regular behavior persists, we need to prove that the successive nonzero values of both F_k and G_k reproduce D_{k-1}. Now look at some data.

$$
\begin{array}{lcccccccccccccccc}
D_1 = & 1 & & -1 & & & & & & & & & & & & & \\
D_2 = & 1 & & 1 & -1 & -1 & & & & & & & & & & & \\
F_2 = & & & 1 & -1 & & & & & & & & & & & & \\
G_2 = & 1 & & & & -1 & & & & & & & & & & & \\
D_3 = & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & & & & & & & & \\
F_3 = & & 1 & 1 & -1 & & -1 & & & & & & & & & & \\
G_3 = & 1 & & & & 1 & & -1 & -1 & & & & & & & & \\
D_4 = & 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\
F_4 = & & 1 & 1 & 1 & -1 & & 1 & -1 & & -1 & & & -1 & & & \\
G_4 = & 1 & & & & & 1 & & & 1 & & -1 & 1 & & -1 & -1 & -1
\end{array}
$$

We see a new regularity: the nonzero sections of F_k all end with a -1, while the nonzero sections of G_k (except for the last one) all end with a $+1$. In fact it seems that D_{k+1} can be obtained from two copies of D_k by the following operation, which we call Rule C (for both Cut and Conway):

Rule C.

(i) cut the first copy of D_k after every -1, and cut the second copy after every $+1$.

(ii) interleave the two copies, starting with the first piece of the second copy.

LEMMA. *For $k \geqslant 1$, Rule C generates (the differences of) Conway's sequence.*

Proof. We have $D_1 = 1 - 1$. We prove by induction that for all $k \geqslant 1$, $D_k(1) = 1$, $D_k(2^k) = -1$, the sum of the elements of D_k is zero, and $D_k(j) = D(2^k + j)$. All this is clearly true for $k = 1$. Suppose it is true for k. Then according to Rule C the first copy of D_k is cut into 2^k pieces, and the second copy into $2^k + 1$ pieces. So the interleaved sequence D' has $D'(1) = 1$, and $D'(2^{k+1}) = -1$. Also the sum of the elements of D' is zero, since this is true of both copies of D_k. Finally we must show that D' and the expanded versions F', G' of the two copies of D_k are exactly D_{k+1}, F_{k+1}, and G_{k+1} respectively. We examine four cases:

$F'(j) = 1$. Here $D' = 1$, and since this is not the last element of a section of F', $F'(j + 1) \neq 0$.

$F'(j) = -1$. Now $D' = -1$, and $F'(j + 1) = 0$.

Similarly for the two cases with $G' \neq 0$.

In all cases, D', F', and G' track D_{k+1}, F_{k+1}, and G_{k+1} exactly. ■

At this stage we have proved that $a(2^k) = 2^{k-1}$. Also it is clear, by an inductive argument, that the cumulative sums of D_k are non-negative, so that the humps of $A(n)$ never go negative. However we have not proved that D_k is skew-symmetric. It is possible to push the following argument through without this, but it is more elegant to derive the skew-symmetry from a beautiful and more general result due to P. Frankl and J. Kahn, (written communication, Sept. 1, 1988) as follows. Start with any vector X in $\{-1, 1\}^n$. Let X^* be the negative reversal of X. Note that $X = X^*$ iff X is skew-symmetric. Cut X after every -1, and cut X^* after every $+1$. Now interleave the pieces to give a sequence Y in $\{-1, 1\}^{2n}$. (It turns out that the numbers of pieces can differ by at most one. If the numbers are equal, we may start with either X or X^*.)

THEOREM. (Frankl and Kahn). $Y = Y^*$.

The proof is by induction on n.

Applying this result to D_k, we see that D_{k+1} is skew-symmetric.

5. Why does Rule L work? We are concerned with D_k, where

$$D_k = C^{k-1}(D_1)$$

with

$$D_1 = 1 - 1$$

and C is the rule given above. We want to prove that for $k \geqslant 3$, $D_{k+2} = L(D_k)$ where L is the substitution rule that we guessed above. Let's look at more data:

$$\begin{array}{ll} D_5 = & 1\,1\,1\,1\,1 - 1\,1\,1 - 1\,1 - 1 - - 1\,1 - 1 - - 1 - - - 1 - - - - - \\ F_5 = & \quad 1\,1\,1\,1 - \quad 1\,1 - \quad 1 - \quad - \quad\quad 1 - \quad - \quad\quad - \quad\quad\quad - \\ G_5 = & 1 \quad\quad\quad 1 \quad\quad 1 \quad\quad 1 \quad - 1 \quad\quad 1 \quad - 1 \quad - - 1 \quad - - - - \end{array}$$

Our L-parsing rules want to view sections of D_5 as corresponding to the elements $3, -1, 1, -3$ of d_3. We mark these divisions:

$$\begin{array}{ll} D_5 = & 1\,1\,1\,1\,1 - 1\,1\,1 - 1\,1 / - 1 - - / 1\,1 - 1 / - - 1 - - - 1 - \quad - - - - \\ F_5 = & \quad 1\,1\,1\,1 - \quad 1\,1 - \quad 1 / - \quad - \quad / \quad 1 - \quad / - \quad\quad - \quad\quad\quad - \\ G_5 = & 1 \quad\quad\quad 1 \quad\quad 1 \quad / \quad 1 \quad - / 1 \quad\quad 1 / \quad - 1 \quad - - 1 \quad - - - - \end{array}$$

which also divide F_5 and G_5 into sections. Since F_5 and G_5 are expanded versions of D_4, the sections give parsings of D_4. To keep the notation specific, we refer to $D_4(F_5)$ and $D_4(G_5)$. Calculations up to the case $k = 8$ show that in all these cases, the sections of $D_{k-1}(F_k)$ that are obtained when D_k is L-parsed according to d_{k-2} can be obtained by the following rule.

Rule M.

(i) replace x (in d_{k-2}) by $m(x)$, where

$$\begin{array}{lllll} m(1) = 1 & -1 & & & \\ m(2) = 2 & -1 & \quad 1 & & \\ m(3) = 3 & -1 & \quad 2 & -1 & \quad 1 \\ m(4) = 4 & -1 & \quad 3 & -1 & \quad 2 \quad -1 \quad\quad 1 \end{array}$$

etc.

$$\begin{aligned} m(-1) &= -2 \\ m(-2) &= -2 \\ m(-3) &= -3 \\ m(-4) &= -4 \end{aligned}$$

etc.

(ii) increase the first element by 1.

Notice that $m(1)$ and $m(-1)$ are anomalous, as with Rule L.

The rule for $D(G)$, called rule $\overline{\text{M}}$, is to replace x by the negative reverse of $m(-x)$, and to reduce the last element by 1.

The M and $\overline{\text{M}}$ rules do not work if we start with an arbitrary skew-symmetric vector, though they do seem to for the special ones that are generated from D_1 (but only from $k = 5$ onwards). We need to show that the M and $\overline{\text{M}}$ rules are consistent with Rule L, and also with Rule C (which defines Conway's sequence).

A one-line explanation of how it comes about that the local rule L works, when Rule C is not local, is that the vectors F and G carry the non-local information, while their reduced forms $D(F)$ and $D(G)$ obey local rules.

THEOREM. *For $k \geqslant 5$, the vectors D_k, F_k, G_k that are generated from D_1 by repeated application of Rule C satisfy the following*:

$$D_k = L(d_{k-2})$$
$$D_{k-1}(F_k) = M(d_{k-2})$$
$$D_{k-1}(G_k) = \overline{M}(d_{k-2}).$$

Proof. Consider how Rule C gets D_k from D_{k-1}. Over the range $2 \leqslant j \leqslant 2^k - 1$, we have 2^{k-2} pieces of the first copy of D_{k-1}, and an equal number of pieces of the second copy. We shall show that over this range, the interleaved pieces are consistent with the first parts (part (i)) of Rules L, M, and $\overline{M}$. Then the second parts of the rules will take care of making $D_k(1) = 1$ (corresponding to an extra section of G_k) and $D_k(2^k) = -1$ (corresponding to extending the last piece of G_k by an additional -1). All we have to do is to check that the sections of $D_{k-1}(F_k)$ and $D_{k-1}(G_k)$ that are generated by Rules M and $\overline{M}$ fit together properly.

First, for $d_{k-2}(j) = \pm 1$, we have

$d_{k-2} = +1$					$d_{k-2}(j) = -1$			
in D_k	1	1	−	1	−	1	−	−
in F_k		1	−	1	−		−	
in G_k	1			1		1		−

while for $d_{k-2}(j) = x \geqslant 2$ the pattern is consistent:

in D_k	1	1^x	−	1	1^{x-1}	−	1	⋯	−	1	1
in F_k		1^x	−		1^{x-1}	−		⋯	−		1
in G_k	1			1			1			1	

and for $d_{k-2}(j) = -y \leqslant -2$

in D_k	−	−	1	−	$-^2$	1	⋯	−	$-^y$
in F_k	−			−			⋯	−	
in G_k			− 1		$-^2$	1	⋯		$-^y$.

We see that in all cases, $D_{k-1}(F_k)$ and $D_{k-1}(G_k)$ do agree with what Rules M and $\overline{M}$ require. The only slight complication is that the influence of each section does extend a little on either side. Thus for $d_{k-2}(j) = 1$ we can fill in some adjacent values:

in D_k	−	/	1	1	−	1	/	−
in F_k		/	1	−		/	−	
in G_k		/	1		1	/		

and similarly in the other cases. But this is O.K., since successive elements of d have alternating signs so these adjacent entries are always consistent with what is needed. ■

6. Computations. Now that we have an efficient way (Rule L) of extrapolating Conway's sequence, we try to use it to answer his two questions. We are interested

in the ratio $a(n)/n$, or equivalently $A(n)/n$, so interesting places are the "first maxima" that occur at

$$\begin{array}{lllllllll} n = & 11 & 23 & 44 & 92 & 178 & 370 & 719 & 1487 \\ A = & 3 & 5 & 8 & 14 & 24 & 48 & 79 & 149 \quad \text{etc.,} \end{array}$$

the "highest foothills" that occur at

$$\begin{array}{llllll} n = & 20 & 82 & 335 & 1361 & \\ A = & 4 & 12 & 39 & 135 & \text{etc.,} \end{array}$$

in the even humps, and at

$$\begin{array}{llllll} n = & 41 & 168 & 684 & & \\ A = & 7 & 22 & 74 & \text{etc.,} & \end{array}$$

in the odd ones, and also the lesser foothills that are visible in the figures. Let us look first at the stretch between the middle of the kth hump and the "first maximum," or equivalently (by symmetry) the stretch between the middle and the "last maximum." Let db_k be the second half of d_k. This generates the right-hand half of the kth hump, from the middle at $n = 3 \cdot 2^{k-1}$ to the upper zero at $n = 2^{k+1}$. We place an asterisk by the term that gives the "last maximum." The vectors are parsed according to Rule L (we need only part (i)).

db_3 1* −3
db_4 1* −2 1 −4
db_5 2 −1 1*/ −2 1 −3 1 −5
db_6 2 −1 1*/ −2 1 −2/ 2 −1 1/ −2 1 −3 1 −4 1 −6
db_7 3 −1 2/ −1 1 −2/ 2 −1 1*/ −2 1 −3/ 2 −1 1/ −2 1 −3 1 −4/...
db_8 same as db_7 this far...
db_9 4 −1 3 −1 2/ −1 1 −2/ 3 −1 2/ −1 1 −2/ 2 −1 1/ −2 1 −2/
3 −1 2/ −1 1 −2/ 2 −1 1*/ −2 1 −3/ ...
db_{10} same as db_9 this far...

We see that the entry in db_k that generates the "last maximum" in the kth hump propagates to the entry in db_{k+2} that generates the "last maximum" in the $k + 2$nd hump. Also, each even hump behaves exactly like the preceding odd hump over this stretch. So all we need look at are the leading parts of these vectors for k odd. Put

$$k = 2m + 1$$

and call the truncated vectors h_m. Keeping the parse marks for clarity,

$h_1 = 1$
$h_2 = 2 -1 \quad 1$
$h_3 = 3 -1 \quad 2/ -1 \quad 1 -1/ 2 -1 \quad 1$
$h_4 = 4 -1 \quad 3 -1 \quad 2/ -1 \quad 1 -2/ 3 -1 \quad 2/ -1 \quad 1 -2/$
$2 -1 \quad 1/ -2 \quad 1 -3/ 3 -1 \quad 2/ -1 \quad 1 -2/ 2 -1 \quad 1$

Notice that these vectors are generated by a completely regular rule; we don't have to worry about the anomalous part (ii) of the full L-rule. What we need are

$$\sum_j h_m(j),$$

which gives the excess of the maximum over $A(3 \cdot 2^{2m})$, and

$$\sum_j |h_m(j)|,$$

which gives the position of the last maximum, measured from $3 \cdot 2^{2m}$.

7. Another breakthrough. We do not need to capture the detailed structure of h_m. All we need is the frequency distribution of its elements. Thus all we need to study are these frequency distributions:

	value of element of h_m								
	-4	-3	-2	-1	1	2	3	4	5
fh_1					1				
fh_2				1	1	1			
fh_3			1	3	2	2	1		
fh_4		1	4	9	6	5	3	1	

which are propagated by the following consequence of Rule L:

an element with value	gives rise to (column headings as above)								
-3	1	1	1		2				
-2		1	1		1				
-1			1	1	1				
1				1	1	1			
2				1		1	1		
3				2		1	1	1	
4				3		1	1	1	1

etc. We can regard this array as an infinite matrix T, so that $fh_{m+1} = fh_m * T$.

Now some numerology leads to formulas for the entries in these vectors of frequencies. If we define

$$S(j, r) = \sum_{k=0}^{k=r} \binom{2k + j}{k}$$

then for $j > 0$

$$\begin{aligned} fh_m(-j) &= S(j - 1, m - j - 1) \\ fh_m(1) &= 1 + S(2, m - 3) \\ fh_m(2) &= 1 + S(1, m - 3) \\ fh_m(j + 2) &= S(j - 1, m - j - 2), \end{aligned}$$

and we can verify that these propagate correctly. Hence we get formulas for the height and position of the "first maximum" in the $2m$ + 1st hump, both relative to the center value:

$$\begin{aligned} \text{height}(m) &= S(0, m - 1) - S(2, m - 2) \\ \text{position}(m) &= S(1, m - 1). \end{aligned}$$

Before we can translate this into the height of the maximum of $A(n)$ itself, we still need to establish the value of $A(3 \cdot 2^{2m})$. So we go through the same kind of analysis, using the full vector da_k (first half of d_k). Now we have to use part (ii) of Rule L, but things turn out simply anyway; the frequency distributions are

	−5	−4	−3	−2	−1	1	2	3	4	5	6	7	8	9
in da_3					1			1						
da_5				1	3	1	1	1		1				
da_7			1	4	11	5	4	3	2	1		1		
da_9		1	5	16	42	22	16	11	7	4	2	1		1

etc. These frequencies turn out to have simple forms:

for $1 \leqslant x \leqslant 2m - 1$, the frequency of x in da_{2m+1} is $\sum_{j=0}^{m-2} \binom{2m-x-1}{j}$.

Also there is a single $x = 2m + 1$.

for $-m \leqslant -y \leqslant -1$, the frequency of $-y$ in da_{2m+1} is $\sum_{j=m-1}^{2m-y-1} \binom{2m-y-1}{j}$

and these propagate correctly. Hence the value of $A(3 \cdot 2^m)$ is indeed $\binom{2m}{m}$ as we guessed above. So finally the "first maximum" in the $2m$ + 1st hump

occurs at $n = 3 \cdot 2^{2m} - S(1, m - 1)$

and has value $A[n] = 1 + S(1, m - 1) + S(0, m - 1)$.

It is easy to see that at this "first maximum" (and indeed at the middle $(3 \cdot 2^{2m})$ we have

$$A(n)/n = O(m^{-1/2}) = O\big((\log n)^{-1/2}\big).$$

Calculations show that $A(n)$ at these "first maxima" decreases in successive humps, and is first below .05 in the 32nd hump. We need to check that the "foothills" have not grown large enough to make $A(n)/n > .05$ in this hump. To expedite the ensuing calculations, let us calculate the progeny of a term x in some d-vector, not at either end. Writing

$$\Delta n = \Sigma^+ + \Sigma^-$$
$$\Delta A = \Sigma^+ - \Sigma^-,$$

the same kind of analysis as above (which was for the case $x = 1$) shows that after p stages of expansion the term x becomes a stretch given by

x	Σ^+	Σ^-
1	$S(0, p)$	$S(2, p - 1)$
2	$1 + S(1, p)$	$S(3, p - 1)$
3	$1 + S(0, p) + S(2, p)$	$S(2, p - 1) + S(4, p - 1)$
4	$1 + S(0, p) + S(1, p) + s(3, p)$	$S(2, p - 1) + S(3, p - 1) + S(5, p - 1)$
5	$1 + S(0, p) + S(1, p) + S(2, p) + S(4, p)$	$S(2, p - 1) + S(3, p - 1) + S(4, p - 1) + S(6, p - 1)$

and so on. For a term $-x$ we simply interchange Σ^+ and Σ^-.

Some values are given in the table.

	1		2		3		4	
$x =$	Δn	ΔA	Δn	ΔA	Δn	Δa	Δn	ΔA
$p = 0$	1	1	2	2	3	3	4	4
1	4	2	6	4	11	7	17	11
2	14	4	21	9	42	18	70	32
3	49	9	77	23	161	51	281	99
4	175	23	287	65	617	155	1112	320

For example, starting with $x = 2$, after two expansions using the first part of Rule L we have

$$4\ -1\ \ 3\ -1\ \ 2/\ -1\ \ 1\ -2/\ \ 3\ -1\ \ 2$$

with $\Delta n = 21$, $\Delta A = 9$. For x negative, simply change the sign of ΔA.

Now it is easy to calculate the progeny of the stretch $-1\ \ 2$ that takes us from the foothill to the first maximum in each of the third and fourth humps. We find that in the 32nd hump the foothill has $A(n)/n > .05$, while in the 33rd hump $A(n)/n$ stays below .05.

What follows is an abbreviated account of a calculation to determine the last n for which $A(n)/n > .05$. This was done essentially by hand, except for using the S language [3] as a glorified desk calculator. Full details are available from the author on request. Start with $da_4 = 4\ -1\ \ 2\ -1$ and label three points p_0, p_1, p_2 as shown. After 14 cycles of expansion by Rule L, these correspond to points P_0, P_1, P_2 on the 32nd hump, whose positions can be calculated as above. We find that at P_2, $A > n/20$, while at P_1, $A < n/20$. We expand the section between p_2 and p_1 in da_4 by Rule L, and look at the corresponding section in da_6. Now we identify new points p_3 and p_4, compute n and A at the corresponding points P_3 and P_4 in the 32nd hump by 13 cycles of expansion, and so on. By the time we get to da_{24} we have pinned down the critical value of n to the stretch between P_{23} at $n = 6083008361$ and P_{24} at $n = 6083010090$. Detailed calculation of this stretch of da_{32} (by fourfold expansion of the stretch $-4\ 3$ in d_{24}) enables us to draw FIGURE 3 and to identify the critical n as 6083008742.

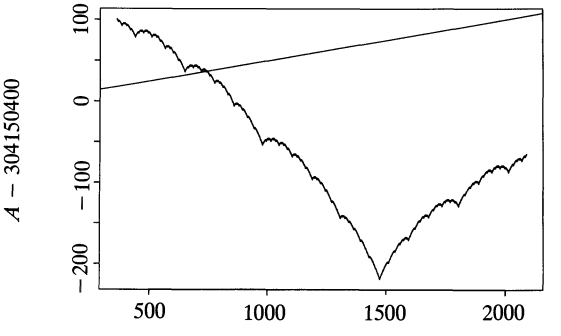
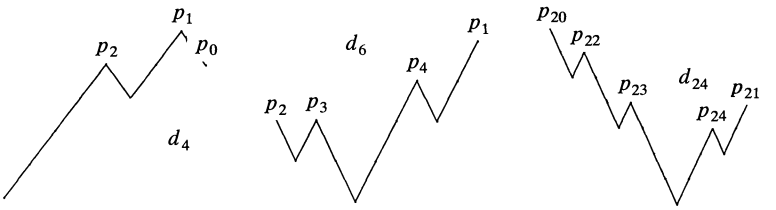


FIG. 3. The last time $A(n)/n$ exceeds 1/20 is at $n = 6083008742$.



8. An aside. When I first tried to compute Conway's number, I neglected to check the foothills of the 32nd hump and so arrived at a point on the 31st hump, with $n = 3173375556$. In this calculation I took only a few steps of the type explained above, and used a Cray computer (not for its speed, but because it has a large memory) to compute 1747771 terms of d_{31} and to search for the correct n. It is amusing to observe that all this effort was unnecessary; neither of these large numbers is the correct response to Conway's challenge. Conway and I both forgot that we were looking at $A(n)$, not $a(n)$; $a(n)/n - 1/2$ is last above $1/20$ at $n = 1{,}489$.

9. Asymptotics. We have

$$da_3 = 1\ -3$$
$$da_5 = 2\ -1\quad 1\ -2\quad 1\ -3\quad 1\ -5 = L(1\ -2)\quad 1\ -5$$

and it is easy to show by induction that for $k = 2m + 1$

$$da_k = L^{m-1}(1 - 2)L^{m-2}(1 - 4) \cdots L^1(1 - (2m - 2))\ 1\ -(2m + 1).$$

Hence the following points are on the $2m + 1$th hump, and the function that interpolates them linearly lies everywhere below the other points of that hump:

$$n_0 = 3 \cdot 2^{2m} = \text{middle} \qquad A_0 = \binom{2m}{m}$$
$$n_1 = n_0 + \binom{2m+1}{m} \qquad A_1 = \binom{2m}{m-1}$$
$$n_2 = n_1 + \cdots \binom{2m+1}{m-1} \qquad A_2 = \binom{2m}{m-2}$$
$$n_m = n_{m-1} + 2m + 1 \qquad A_m = 1.$$

As $m \to \infty$, the points $(n_j 2^{-2m}, A_j(m/2)^{1/2}2^{-2m})$ for $j = 0, 1, \ldots, m$ converge to the curve with parametric definition

$$x = 2 + 2\Phi(u)$$
$$y = \phi(u) \qquad -\infty < u < \infty$$

where ϕ, Φ are the standard Gaussian density and cumulative functions respectively. FIGURE 4 shows this curve. The extreme tangent has slope .138015 (corresponding to $u = -.27603$) and occurs at $x = 2.78253$.

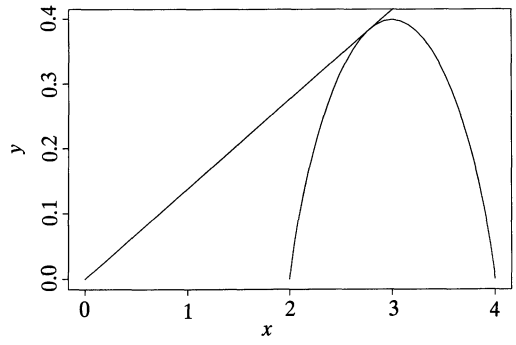


FIG. 4. The asymptotic shape of the humps in $A(n)$.

What happens between these "interior" points? FIGURE 5 shows what happens if we divide the kth hump (suitably scaled) by the asymptotic function, for $k = 3, \ldots, 12$. It seems that the ratio may be tending to unity for all x in $(2, 4)$, but I cannot prove this. If this conjecture is correct, an estimate of the maximum value of $a(n)/n - 1/2$ for $2 \cdot 2^{2m} < n < 4 \cdot 2^{2m}$ is $.09759m^{-1/2}$. An estimate of n_ε, the largest n for which $A(n)/n > \varepsilon$, is $5.565 \cdot 2^{2m}$ where m is the largest integer in $(.346/\varepsilon)^2/\pi$.

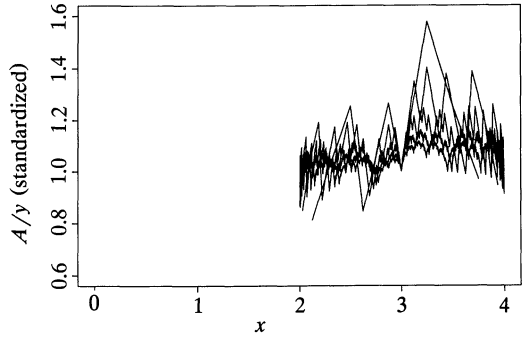


FIG. 5. Sections of $A(n)/y$, standardized.

We can get a rigorous bound by noting that in the kth hump, $A(n)/n$ cannot exceed $M_k 2^{-k}$ where M_k is the maximum of A in the kth hump, which we found above. Numerically,

epsilon	bound	estimate
.01	$7.5 \cdot 10^{1104}$	$1.2 \cdot 10^{580}$
.02	$8.9 \cdot 10^{277}$	$2.5 \cdot 10^{145}$

10. Some further comments. D. R. Hofstadter (private communication, Sept. 2, 1988) has found some ancient notes in which he studied Conway's sequence, and derived the following rule for its behavior. Let

$$f(j) = \#\{n: a(n) = j\} \qquad j = 1, 2, 3, \ldots .$$

Breaking this sequence at powers of 2, we find

2
2
1 3
1 1 2 4
1 1 1 2 1 2 3 5
1 1 1 1 2 1 1 2 1 2 3 1 2 3 4 6

and there is a simple (local) rule for generating each section from the previous one; in fact Hofstadter found two such rules.

(i) Reading the previous row backwards, replace any integer $k > 1$ as a group of $k - 1$ 1's followed by a 2,
Replace any group of k 1's by the group 3 4 5 ... $k + 2$.
Add 1 to the last number you produce.

(ii) Reading the previous row forwards, replace k by 1 2 3 ... $k - 1$ k, except when k is the rightmost number of its line, in which case replace it by 1 2 3 ... $k - 1$ $k + 1$.

I do not see how to prove that these rules work, except by an analysis similar to mine. (The details for (ii) are rather simpler than for my rules. For (i) we seem to need the symmetry property.)

There is even a formula for $a(n)$, due to G. Phillips, shown to me by Conway on Sept. 9, 1988. Consider the following array.

					1	1
				2	1	1
			4	1	2	1
		8	1	3	3	1
	16	1	4	6	4	1
32	1	5	10	10	5	1

etc. Given any n between 2^k and 2^{k+1}, find the row that starts with 2^k, and mark as many successive elements as possible without letting the sum of the marked elements exceed n. When the next entry in that row would take the sum above n, move up one row and continue. At most one element in each column may be marked. Stop when the total of the marked elements is n. For example, when $n = 23$ we mark the elements 16, 1, 4 in the fifth row, then the first 1 in the third row, and finally the first 1 in the second row.

Now $a(n)$ can be read off as the sum of the elements that are north-east of each of the marked elements. For $n = 23$, these are 8, 1, 3, 1, 1, so that $a(23) = 14$.

11. Some related sequences, and further comments. Conway's sequence is anomalous at $n = 2$. Applying his recursion starting at $n = 2$ (with $a(1) = 1$) gives a very dull sequence. What if we start in other ways? We have to be careful that $a(n)$ stays below n, or the sequence will not be well-defined. An interesting class is obtained by starting with a string of m 1's, applying Conway's recursion for $n \geqslant m + 1$. For $m = 3$, we get a sequence that proceeds in stretches between the Fibonacci numbers, and that has structure very similar to that found for Conway's sequence. Here we have to interleave cut copies of D_{k-2} and D_{k-1}. Rule L (with a modified part (ii)) generates D_k from D_{k-3}. Similarly for all $m \geqslant 4$; we leave

the details as an exercise for the reader. (On 1/4/89 R. K. Guy pointed out (personal communication) that this is problem E3274 [5]!)

Another variation is obtained by writing

$$b(n) = b(b(n-1)) + b(n-1-b(n-1)),$$

but this turns out to be a trivial modification of $a(n)$; in fact $b(n-1) = n - a(n)$.

In his delightful book [4], D. R. Hofstadter mentions the (nonmonotone) sequence generated by

$$q(n) = q(n - q(n-1)) + q(n - q(n-2))$$

which has bizarre behavior. The first 1000 values exhibit some striking regularities, but these do not persist; FIGURE 6 shows the first 20,000 values. There is obvious approximate structure, but I have no idea what happens asymptotically.

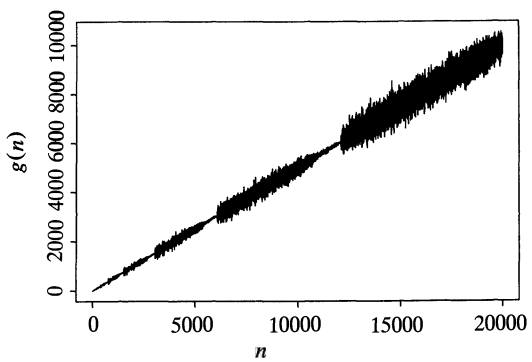


FIG. 6. Hofstadter's q-sequence.

Finally, a minor change in Conway's definition gives the monotone sequence

$$c(n) = c(c(n-2)) + c(n - c(n-2)).$$

Now the increments are very irregular; see FIGURE 7. It is not clear whether $c(n)/n$ tends to a limit.

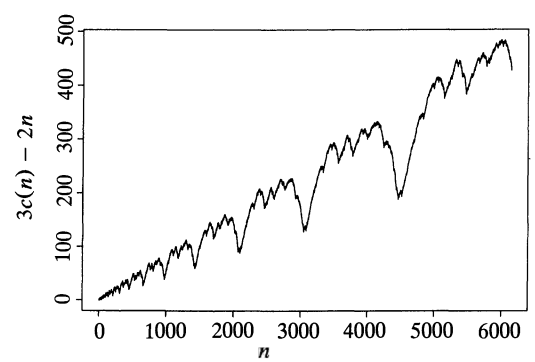


FIG. 7. $3c(n) - 2n$ for $1 \leq n \leq 6144$.

REFERENCES

1. J. H. Conway, Some crazy sequences, videotaped talk at AT&T Bell Labs, July 15, 1988.
2. C. L. Mallows and J. W. Tukey, An overview of techniques of data analysis, in Some Recent Advances in Statistics, ed. de Oliviera and Epstein, Academic Press, London, 1982.
3. R. A. Becker, J. M. Chambers, and A. R. Wilks, The New S Language, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1988.
4. D. R. Hofstadter, Gödel, Escher, Bach, Vintage Books, New York, 1980, p. 137.
5. David Newman, Problem E3274, *Amer. Math. Monthly*, 95 (1988) 555.

Notice to authors: Recently the *Monthly* has not been processing papers because of a very high backlog. This situation has now eased somewhat and we are once again considering new manuscripts for possible publication. Thank you for your patience.—The Editors.

LETTERS TO THE EDITOR

To the Editor:

In a recent MONTHLY article [3], the authors give a Taylor polynomial-like formula to define $f(A)$, where A is a square complex matrix and $f(\cdot)$ is a scalar-valued (real or complex) function of a scalar. Although the authors assert that their formula (3.1-2) for $f(A)$ "has not been given previously in this form", it was published by H. Schwerdtfeger in 1938; see Section 6 in [5]. For an application of Schwerdtfeger's formula to calculating $(d/dt)(f(A(t))$, see [1] and Section 6.6 of [2]. For a discussion of several equivalent definitions of $f(A)$, including Schwerdtfeger's formula, see [4] as well as Sections 6.1-2 of [2].

Sincerely yours,

Roger A. Horn
Johns Hopkins University

[1] Ju. L. Daleckii, Differentiation of Non-Hermitian Matrix Functions Depending on a Parameter, American Mathematical Society Translations, Series 2, Vol. 47, American Mathematical Society, Providence, Rhode Island, 1965, 1–30. Russian original in *Voronez. Gos. Univ. Trudy Sem. Funkcional. Anal.* No. 1 (1956), 81–105.

[2] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, New York, 1990.

[3] R. R. London and H. P. Rogosinski, Decomposition Theory in the Teaching of Elementary Linear Algebra, this MONTHLY, 97 (1990), 478–485.

[4] R. F. Rinehart, The Equivalence of Definitions of a Matric Function, this MONTHLY, 62 (1955), 395–413.

[5] H. Schwerdtfeger, Les Fonctions de Matrices I, Les Fonctions Univalentes, *Actualités Scientifiques et Industrielles*, No. 649, Paris, Hermann, 1938.

To the Editor:

I would like to correct an important detail in Richard Guy's review of my book *The Mathematics of Plato's Academy, A New Reconstruction* (this MONTHLY, May 1990, 410–413): the book **does** contain, on pages 356–360, a description of Bill Gosper's algorithm for continued fraction arithmetic, together with a tableau illustrating its use in evaluating $\sqrt{2} \times \sqrt{3}$. The description ends with an implied challenge to give a direct proof that, in this case, the output will be periodic; and a much more significant contribution would be to go on and see if it is possible to use the algorithm to prove that the output from some example like $\sqrt{2} + \sqrt{3}$ is, or is not, bounded.

Yours sincerely,

Daniel Fowles
University of Warwick
Coventry, UK

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

A DYNAMIC PUZZLE

MOSHE ROSENFELD
Department of Mathematics and Computer Science, Pacific Lutheran University, Tacoma, WA 98447

A dynamic puzzle: Start with an $n \times m$ rectangle partitioned into nm squares.

Step 0: Select any square and cover it.

Step $k + 1$: Cover a set of uncovered squares by a translate of the current covered set.

Objective: Cover as many squares as possible.

Observe that each step doubles the number of covered squares, so that after the k-th step exactly 2^k squares are covered. In FIGURE 1 below we show a covering of 128 squares of a 14×14 square (a *cover of size* 128, in the rest of this note the term cover will be used only to describe sets covered using the above steps). We invite the reader to try to cover as many squares as possible on a 7×7 square (32?), on a 23×23 square (512?) or on a 47×47 square (2048?).

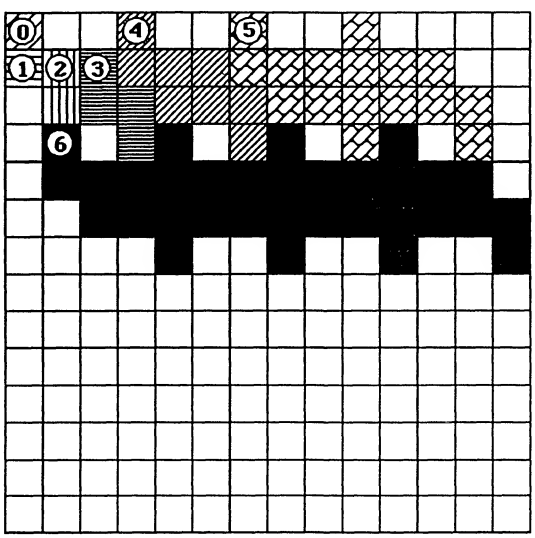


FIG. 1. A 14×14 square with a cover of size 128. The numbers inside the squares mark the image of the first square (0) when the k-th translation is executed. The figure shows a cover of size 64 for the 7×14 rectangle. There are two distinct ways to obtain a cover of size 128. One cover of size 128 is obtained by placing a copy of the cover in the 7×14 bottom rectangle, the second can be obtained by translating the current cover down to the 6-th row.

The term puzzle is frequently used to describe a game or problem where one is asked to put pieces together to form a given shape. In this sense this puzzle can be viewed as a dynamic puzzle, the pieces to be fitted into the initial rectangle change dynamically. Every step determines the shape of the subsequent piece to be placed in the rectangle. For a given rectangle, we are interested in finding the largest number of squares that can be covered and the number of distinct shapes of the final covered set. Clearly, one cannot cover more than 128 squares of a 14×14 square. From FIGURE 1 one can immediately see that there are at least two distinct final shapes, in fact, FIGURE 1 also indicates a cover of size 128 for the 14×13 rectangle. While the rectangle is a "simple" object, it is not so nice "mathematically", not all squares are "created equal" even if the initial rectangle is a square. We can do our dynamic puzzle on a "nicer" object, the torus. One advantage of using the torus is the choice of the first square. Since all squares are "created equal," we can choose the first square as we wish. Clearly, if an $n \times k$ rectangle has a cover of size 2^m then so does the $n \times k$ torus, the reverse is not true. The 6×6 torus has a cover of size 32 while the 6×6 square does not. Two covers are *distinct* if one is not congruent to the other. In FIGURE 2 we show two distinct covers of size 32 of the 7×7 torus (we invite the reader to determine whether the 7×7 square admits a cover of size 32).

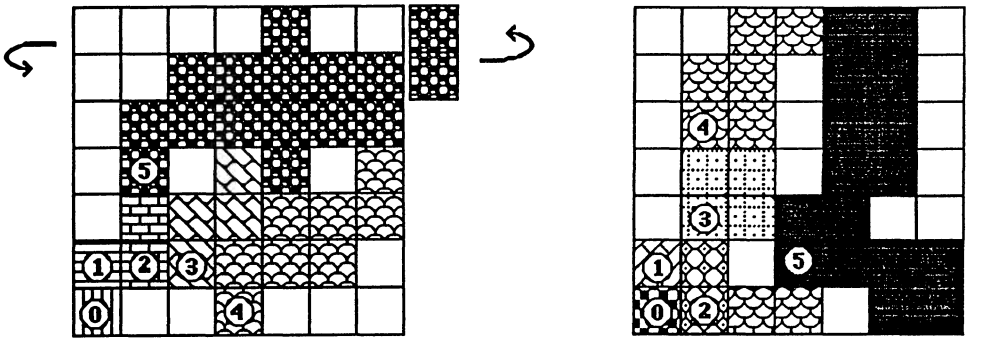


FIG. 2. Two distinct covers of size 32 of the 7×7 torus (actually cylinder).

The puzzle on the torus exhibits some unexpected phenomena. Consider for example the three tori of sizes 5×14, 6×12 and 5×15. They have 70, 72 and 75 squares respectively. FIGURE 3 shows covers of size 64 for the first two (the "smaller") tori. Surprisingly, it is not possible to cover 64 squares of the "larger" 5×15 torus (a proof is included in [**1**]).

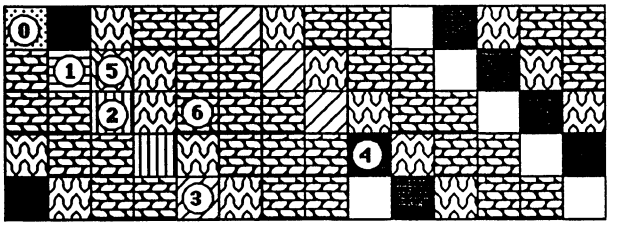


FIG. 3. A cover of size 64 of the 5×14 torus

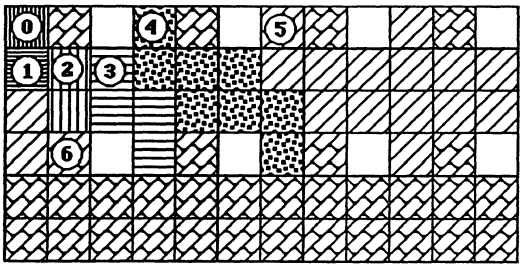


FIG. 4. A cover of size 64 of the 6×12 torus

This puzzle has a natural generalization to higher dimensional boxes and tori, exhibiting the same "paradoxical" behavior. For example we can show that the $7 \times 7 \times 7$ torus (with 343 cubes) does not admit a cover of size 256 but the "smaller" 3-dimensional $6 \times 7 \times 8$ (only 336 cubes) does admit a cover of size 256.

While the smaller size problems can be solved by backtracking on a computer, the backtracking becomes intractable very quickly. We used backtracking to prove the nonexistence of a cover of size 64 for the 5×15 torus but only after finding a substantial reduction in the number of possibilities. Most of the results known to us are for the 2-dimensional puzzle on the torus and the square. The following is a sample of our results. If n and k are relatively prime then the $n \times k$ torus has a cover of size $2^{[\log_2 nk]}$ (largest possible cover). If $3*2^{n-1} \leqslant k < 2^{n+1}$ then it is possible to cover 2^{2n+1} squares of a $k \times k$ torus. We have similar results for the $k \times k$ square.

The following is a list of related interesting problems.

1. The smallest square torus for which we do not know the size of the largest cover is the 47×47 torus. More generally, for all sizes $k \geqslant 47$, for which $2^n\sqrt{2} < k < 3*2^{n-1}$ we do not know whether it is possible to cover 2^{2n+1} squares of the $k \times k$ torus (or square).
2. For any pair of integers n and m, determine the largest possible cover of the $n \times m$ torus (or rectangle). For the answer in case $\gcd(m, n) = 1$ see [1].
3. For any given sizes, determine the number of distinct final maximal covers.
4. The higher dimensional puzzle seems much more complicated than the 2-dimensional. It will be interesting to obtain bounds for such tori (or boxes).
5. Sherman Stein suggested the following interesting variation of this puzzle: instead of starting with a single square in step 0, start with some initial set of squares, for example start with three squares in a row or L-shaped and then continue to duplicate the covered set. Again the objective is to cover the largest possible number of squares.

REFERENCE

1 B. Dorner, C. Meyer and M. Rosenfeld, A square puzzle (in final stages of preparation, preprints available).

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A Continued Fraction and Permutations With Fixed Points

HENRI DARMON
Department of Mathematics, Harvard University, Cambridge, MA 02138

JOHN MCKAY
Department of Computer Science, Concordia Univ., Montreal, Canada H3G 1M8

The set Σ_n of all permutations on $\Omega_n = \{1, 2, 3, \ldots, n\}$ is partitioned into FPF_n permutations acting without fixed points and FP_n acting with at least one fixed point. These numbers are related by

$$|\Sigma_n| = n! = FPF_n + FP_n \quad \text{for} \quad n \geq 0,$$

with the convention that $FPF_0 = 1$ and $FP_0 = 0$. We shall index all sequences from 0.

Let V denote the two-dimensional real vector space of sequences $\{t_n\}$ satisfying the recurrence

$$t_{n+1} = n(t_n + t_{n-1}).$$

To convince the reader that V is worthy of study, consider the following:

THEOREM. *The sequences* $\{n!\}$, $\{FPF_n\}$, *and* $\{FP_n\}$ *belong to* V.

Proof. The proof that $\{n!\} \in V$ is trivial. To show that $\{FPF_n\} \in V$ we observe that any fixed-point-free permutation in Σ_{n+1} can be obtained from the identity permutation in exactly one of two ways:

(1) by applying a permutation with a unique fixed point $i \in \Sigma_n$ to the first n elements, and then transposing $n + 1$ and i, or:
(2) by applying a fixed-point-free permutation to the first n elements, followed by a transposition of $n + 1$ and some $i \in \Omega_n$.

There are n possible choices for i, FPF_{n-1} possible permutations on $1, \ldots, n$ which leave only i fixed, and FPF_n fixed-point-free permutations on the first n elements. So $FPF_{n+1} = n(FPF_n + FPF_{n-1})$, and hence $\{FPF_n\} \in V$. The last statement of the theorem follows from the identity $FP_n + FPF_n = n!$. ■

The vector space V also contains a sequence which converges to 0, which means that the null sequences form a proper (one-dimensional) subspace of V. This is the content of the next theorem:

MAIN THEOREM. *The sequence* $\{I_n\}$ *given by*

$$I_n = \int_0^\infty (e^{-x} - 1)^n e^{-e^{-x}-x}\, dx = \int_0^1 (-t)^n e^{t-1}\, dt$$

belongs to V. *Moreover,* $I_0 = 1 - 1/e$, $I_1 = -1/e$, $\lim_{n\to\infty} I_n = 0$, *and* $\forall n$, $|I_n| < 1$.

Proof. That $\{I_n\}$ satisfies the appropriate recurrence follows from integration by parts. It is a trivial exercise to evaluate I_0 and I_1. The limit of $\{I_n\}$, as well as the bound, follow from dominance of I_n by $\int_0^\infty e^{-x}\,dx$. ■

Let us denote by $S(a, b)$ the sequence $\{t_n\} \in V$ satisfying the initial conditions $t_0 = a$, $t_1 = b$. The previous results can be summarized as follows:

$$S(0,1) = \{FP_n\},$$
$$S(1,0) = \{FPF_n\},$$
$$S(1,1) = \{n!\} = \{FP_n + FPF_n\},$$
$$S\left(1 - \frac{1}{e}, -\frac{1}{e}\right) = \{I_n\}.$$

Choosing an appropriate basis for V, one has

$$S(0,1) = \left(1 - \frac{1}{e}\right)S(1,1) - S\left(1 - \frac{1}{e}, -\frac{1}{e}\right),$$

which leads to the following result:

$$FP_n = \left(1 - \frac{1}{e}\right)n! - \int_0^1 t^n e^{t-1}\,dt. \qquad (*)$$

The approximation $FP_n \approx (1 - 1/e)n!$ is well known in elementary combinatorial probability, where the result is usually stated as:

THEOREM. *After a random shuffle of a deck of cards, the chances that at least one card will remain fixed is very close to* $1 - 1/e$.

Formula $(*)$ provides a closed form expression for the error term.

Our investigation concludes with an evaluation of a continued fraction which is closely related to the previous formula:

THEOREM.

$$\frac{1}{e-1} = \cfrac{1}{1 + \cfrac{2}{2 + \cfrac{3}{3 + \cdots}}}.$$

Proof. The right-hand side converges and is equal to $\lim_{n\to\infty} p_n/q_n$, where $\{p_n\}$, $\{q_n\} \in V$, and $p_0 = q_1 = 1$, $p_1 = q_0 = 0$. So

$$\cfrac{1}{1 + \cfrac{2}{2 + \cdots}} = \lim_{n\to\infty} \frac{S(1,0)}{S(0,1)} = \frac{1/e}{1 - 1/e} = \frac{1}{e-1}. \quad ■$$

The theory of the recurrence relations determining the p_n, q_n appears in [**1**, p. 492], [**2**, p. 15]. An evaluation of the continued fraction is given as an exercise in [**1**, Ch. XXXIV, Ex. 24, p. 576], but it is arrived at quite differently, using a classical formula of Gauss which gives continued fraction expansions of ratios of hypergeometric series.

REFERENCES

1. G. Chrystal, Textbook of Algebra, Part 2, Dover Publications, Inc., New York, 1961 (orig. publ. 1904).
2. H. S. Wall, Analytic Theory of Continued Fractions, Chelsea, New York, 1967 (orig. publ. 1948).

Smoothness From Finite Points

VICTOR NEVES
University of Iowa, Mathematics, Iowa City, IA 52242
UBI, R. Marquês D'Avila e Bolama, 6200 Covilhã, Portugal

1. Introduction. Like other branches of mathematics, infinitesimal analysis allows some informality; but careless use of terms may lead to wrong answers.

Unfortunately, there is a serious error in the definition of tangent plane given in [2]. It is closely related to the well-known "Schwarz Accordion," which is treated in [4, p. 124] (with incorrect algebra). There also seems to be a tendency to blur the distinction between "standard" and "finite." Nevertheless, the main result of [2, p. 434], can be reclaimed with correct definitions. That theorem is actually proved, if one replaces the definition of tangent plane as in section 3, below. Moreover, the corrected proof is certainly simpler then the classical one we first learned from [1], for it goes more directly to the point.

Ours is not the only possible correction to Henle's definition of tangent plane. It suffices to require that the three close points that define the nearly tangent planes, form a triangle with no infinitesimal angles.

We use the notation set up in [3] chapters 1 and 1*. $\mathbb{R}^*$ denotes the nonstandard extension of the set of real numbers $\mathbb{R}$.

We consider $\mathbb{R}$ a subset of $\mathbb{R}^*$. ${}^n\mathbb{R}^* = (\mathbb{R}^*)^n$ $(n = 2, 3)$.

An element r of $\mathbb{R}^*$ is standard, finite, or infinitesimal if, respectively, $r \in \mathbb{R}$, $|r| < s$ for some s in $\mathbb{R}^+$, or $|r| < s$ for all $s \in \mathbb{R}^+$. An element of ${}^n\mathbb{R}^*$, $(n = 1, 2, 3)$, is *standard*, *finite*, or *infinitesimal* if all its coordinates are, respectively, standard, finite, or infinitesimal. We say x is infinitely close to I in ${}^n\mathbb{R}^*$, $x \approx y$, if $x - y$ is infinitesimal.

Also, for $n = 1, 2, 3$, if $a \in {}^n\mathbb{R}^*$ is finite and b is the (unique) standard element of ${}^n\mathbb{R}^*$ which is infinitely close to a, we say b is the *standard part* of a and write $b = \mathrm{st}a$.

Finally, functions and their nonstandard extensions will be denoted with the same symbol.

2. A critique of [2]. Defining closeness of planes with the difference in normal unit vectors is dangerous: a plane is not close to itself if one takes opposite normals. But let us assume that a choice of unit normals was implicitly assumed.

According to Henle's definition [2], a tangent plane to a surface S, at a point u of S, is infinitely close to any plane determined by a three-tuple (u, v, w) of infinitely close, noncollinear points on the surface. This is simply false. This

definition would mean that spheres and cylinders have no tangent planes. For example, there are planes containing the normal to a sphere, or a cylinder, which intersect the surface in a circle. Three distinct points on a circle (infinitely close or not) determine the plane of that circle. In this case three infinitely close, non-collinear points determine a plane perpendicular to the tangent plane.

Contrary to what Henle states, one cannot characterize differentiability of a function, solely with the existence of tangent planes to its graph: when $f:\mathbb{R}^2 \to \mathbb{R}$ is given by $f(x_1, x_2) = x_1^{1/3}$, for all (x_1, x_2) in $\mathbb{R}^2$, the graph of f has a (vertical) tangent plane at $(0,0,0)$ and f is not differentiable at $(0,0)$.

By avoiding reference to $\mathbb{R}^{n*}$ $(n = 1,2,3)$, it is not made clear if the points (a, b) in [2] are supposed to be standard or not. The difference is illustrated in Theorems 3.1 and 3.2, below. The point of those theorems is that, except for functions of class C^1, derivatives at nonstandard finite points are not necessarily infinitesimal approximations to all ratios of infinitesimal increments. If $f(x) = x^2 \sin(2\pi/x)$ for $x \neq 0$ in $\mathbb{R}$, and $f(0) = 0$, then for an infinite hyperinteger n, one has

$$\frac{4}{4n+1} \approx \frac{1}{n} \cdot \frac{f[4/(4n+1)] - f(1/n)}{4/(4n+1) - 1/n} \approx -4 \quad \text{but } f'\left(\frac{1}{n}\right) = -2\pi.$$

3. Calculus. The difference between standard and finite. As we still want to give a geometrical condition of differentiability in terms of tangent planes, some definitions are in order.

We will always give a plane P contained in $\mathbb{R}^{3*}$ by one of its points, a, and a unit vector $u = (u_1, u_2, u_3) \in \mathbb{R}^{3*}$ according to the following equalities

$$P = P(u, a) = \{x \in \mathbb{R}^{3*} : (x - a) \cdot u = 0\}.$$

We take the plane $P((0,0,1),(0,0,0))$ to be "the ground" and say that a plane $P(u, a)$ is *S-vertical* if $u_3 \approx 0$. A standard plane is vertical, $u_3 = 0$, iff it is S-vertical, $u_3 \approx 0$.

Clearly a unit vector is finite. Also, if a is finite in $\mathbb{R}^{3*}$ every finite point in $P(u, a)$ is infinitely close to a standard point in $P(\mathrm{st}u, \mathrm{st}a)$, and this standard plane is vertical iff $P(u, a)$ is S-vertical.

Let $f:\mathbb{R}^2 \to \mathbb{R}$ be a function and let $a = (a_1, a_2, f(a_1, a_2))$. We say the plane $P(u, a)$ in $\mathbb{R}^{3*}$ is *S-tangent* to the graph of f over (a_1, a_2) if for all (b_1, b_2) infinitely close to (a_1, a_2),

$$\frac{b-a}{\|b-a\|} \cdot u \approx 0 \quad \text{holds with} \quad b = (b_1, b_2, f(b_1, b_2)).$$

Note that a standard plane will be tangent to a graph at the point $(a_1, a_2, f(a_1, a_2)) \in \mathbb{R}^3$ iff it is S-tangent over (a_1, a_2).

From now on denote $\Delta f = f(a_1 + \delta, a_2 + \varepsilon) - f(a_1, a_2)$.

3.1. THEOREM. *Let $f:\mathbb{R}^2 \to \mathbb{R}$ be a standard function. The following conditions are equivalent.*

i) *f is differentiable.*

ii) *For all standard $(a_1, a_2) \in \mathbb{R}^2$, there are standard real numbers α and β, such that, when δ and ε are infinitesimal in $\mathbb{R}^*$*

$$\Delta f = \alpha\delta + \beta\varepsilon + \eta\|(\delta, \varepsilon)\|, \quad \textit{for some } \eta \approx 0.$$

iii) *The graph of f has a non-S-vertical, S-tangent plane over all standard* (a_1, a_2) *in* $\mathbb{R}^2$.

This is only a correction to Henle's remark on differentiability [2, p. 434, line 5]. Its proof is indeed an application of standard infinitesimal techniques.

Consider the following result obtained from Theorem 3.1, by replacing "standard" by "finite":

3.2. THEOREM. *Let* $f: \mathbb{R}^2 \to \mathbb{R}$ *be a standard function. The following conditions are equivalent.*

i) *f is of class* C^1.

ii) *For all finite* $(a_1, a_2) \in \mathbb{R}^{2*}$, *there are finite numbers* α *and* β, *such that when* δ *and* ε *are infinitesimal in* $\mathbb{R}^*$ *then*

$$\Delta f = \alpha\delta + \beta + \eta\|(\delta, \varepsilon)\| \quad \textit{for some } \eta \approx 0. \tag{1}$$

iii) *The graph of f has a non-S-vertical, S-tangent plane over all finite* (a_1, a_2) *in* $\mathbb{R}^{2*}$.

This result is similar to 5.7.9. in [4], with $k = 1$.

Proof. i) $\Rightarrow$ *ii).* Assume i). Let (a_1, a_2) be finite, δ and ε be infinitesimal. Using the $*$-transfer of the Mean Value Theorem in $\mathbb{R}^*$ one obtains

$$\begin{aligned}\Delta f &= f(a_1 + \delta, a_2 + \varepsilon) - f(a_1, a_2 + \varepsilon) + f(a_1, a_2 + \varepsilon) - f(a_1, a_2) \\ &= \frac{\partial f}{\partial x_1}(a_1 + \lambda\delta, a_2 + \varepsilon)\delta + \frac{\delta f}{\delta x_2}(a_1, a_2 + \mu\varepsilon)\varepsilon \quad \text{for some } \lambda, \mu \in [0, 1].\end{aligned}$$

By continuity of the partial derivatives, it follows that

$$\Delta f = \left[\frac{\partial f}{\partial x_1}(a_1, a_2) + \eta_1\right]\delta + \left[\frac{\partial f}{\partial x_1}(a_1, a_2) + \eta_2\right]\varepsilon \quad \text{with } \eta_1 \approx \eta_2 \approx 0.$$

Finally (1) holds with

$$\alpha = \frac{\partial f}{\partial x_1}(a_1, a_2), \quad \beta = \frac{\partial f}{\partial x_2}(a_1, a_2), \quad \eta = \eta_1 \cdot \frac{\delta}{\|(\delta, \varepsilon)\|} + \eta_2 \cdot \frac{\varepsilon}{\|(\delta, \varepsilon)\|}.$$

ii) $\Rightarrow$ *i).* Assume ii). $\alpha \approx \alpha'$ and $\beta \approx \beta'$ imply

$$\Delta f = \alpha'\delta + \beta'\varepsilon + \eta'\|(\delta, \varepsilon)\| \quad \text{with } \eta' \approx 0 \text{ if } \delta \approx 0 \approx \varepsilon.$$

It follows that, when (a_1, a_2) is standard, we have

$$\Delta f = (\mathrm{st}\alpha)\delta + (\mathrm{st}\beta)\varepsilon + \eta_1\|(\delta, \varepsilon)\| \quad \text{with } \eta_1 \approx 0 \text{ if } \delta \approx \varepsilon \approx 0.$$

Thus, according to Theorem 3.1., f is differentiable and

$$\frac{\partial f}{\partial x_1}(a_1, a_2) = \mathrm{st}\alpha, \qquad \frac{\partial f}{x_2}(a_1, a_2) = \mathrm{st}\beta,$$

when (a_1, a_2) is standard. Next, by $*$-transfer of the classical definition of partial derivative, when (a_1, a_2) is finite and δ is *sufficiently* small but not zero, by ii) we

have

$$\frac{\partial f}{\partial x_1}(a_1, a_2) \approx \frac{1}{\delta}[f(a_1 + \delta, a_2) - f(a_1, a_2)]$$
$$= \frac{1}{\delta}[\alpha\delta + \eta|\delta|] \quad \text{with } \eta \approx 0$$
$$= \alpha + \eta' \quad \text{with } \eta' \approx 0.$$

Thus, we may assume

$$\alpha = \frac{\partial f}{\partial x_1}(a_1, a_2), \qquad \beta = \frac{\partial f}{\partial x_2}(a_1, a_2) \quad \text{for all finite } (a_1, a_2) \in \mathbb{R}^{2*}.$$

We will show that *for all finite* $(a_1, a_2) \in \mathbb{R}^{2*}$ and infinitesimal $\delta \in \mathbb{R}^*$ the following hold

a) $$\alpha(a_1 + \delta, a_2) \approx \alpha(a_1, a_2) \approx \alpha(a_1, a_2 + \delta)$$

b) $$\beta(a_1 + \delta, a_2) \approx \beta(a_1, a_2) \approx \beta(a_1, a_2 + \delta)$$

so that, for $\delta \approx \varepsilon \approx 0$

$$\alpha(a_1 + \delta, a_2 + \varepsilon) \approx \alpha(a_1, a_2 + \varepsilon) \approx \alpha(a_1, a_2)$$
$$\beta(a_1 + \delta, a_2 + \varepsilon) \approx \beta(a_1, a_2 + \varepsilon) \approx \beta(a_1, a_2).$$

In particular, these conditions hold with $(a_1, a_2) \in \mathbb{R}^2$ and the partial derivatives are continuous.

We shall only prove a), since b) can be deduced analogously. Let (a_1, a_2) be finite. Expand Δf three different ways:

$$\Delta f = \alpha(a_1, a_2)\delta + \beta(a_1, a_2)\varepsilon + \eta\|(\delta, \varepsilon)\| \tag{1}$$
$$\Delta f = \alpha(a_1 + \delta, a_2 + \varepsilon)\delta + \beta(a_1 + \delta, a_2 + \varepsilon)\varepsilon + \eta_1\|(\delta, \varepsilon)\| \tag{2}$$
$$\begin{aligned}\Delta f &= f(a_1 + \delta, a_2 + \varepsilon) - f(a_1, a_2 + \varepsilon) + f(a_1 + a_2 + \varepsilon) - f(a_1, a_2) \\ &= \alpha(a_1, a_2 + \varepsilon)\delta + \eta_2|\delta| + \beta(a_1, a_2) + \eta_3|\varepsilon|.\end{aligned} \tag{3}$$

By assumption, if $\delta \approx \varepsilon \approx 0$ then $\eta \approx \eta_i \approx 0$ $(i = 1, 2, 3)$, so combining 1) and 2), with $\varepsilon = 0$, and dividing by δ, $\alpha(a_1 + \delta, a_2) \approx \alpha(a_1, a_2)$ follows for all finite a_1, a_2 and all nonzero infinitesimal δ.

Next combine 1) and 3) and take $\delta = \varepsilon$ in order to obtain $\alpha(a_1, a_2 + \delta) \approx \alpha(a_1, a_2)$ for all nonzero infinitesimal δ.

ii) $\Rightarrow$ *iii)*. Assume ii) and take

$$u = \frac{(\alpha, \beta, -1)}{\|(\alpha, \beta, -1)\|}.$$

$P(u, (a_1, a_2, f(a_1, a_2))$ is S-tangent to the graph of f over (a_1, a_2) and, of course, $-1/\|(\alpha, \beta, -1)\|$ is not infinitesimal, for α and β are finite, so the plane is not nearly vertical.

iii) $\Rightarrow$ *ii)*. Assume iii). Let $\dfrac{(\alpha, \beta, -1)}{\|(\alpha, \beta, -1)\|}$ be a unit vector which defines a non-S-vertical plane, S-tangent to the graph of f over the point $(a_1, a_2) \in \mathbb{R}^{*2}$.

S-tangency yields

$$\alpha\delta + \beta\varepsilon - \Delta f = \mu\|(\delta, \varepsilon, \Delta f)\|, \qquad \text{with } \mu \approx 0. \tag{4}$$

Moreover, α and β are finite since the plane is not S-vertical. Define $\rho = \Delta f/\|(\delta, \varepsilon)\|$ so that $\|(\delta, \varepsilon, \Delta f)\| = \sqrt{1 + \rho^2}\,\|(\delta, \varepsilon)\|$. Dividing by $\|(\delta, \varepsilon)\|$ in equation (4) one obtains $\rho = \|(\alpha, \beta)\|\cos(\theta) - \mu\sqrt{1 + \rho^2}$ (with α and β finite and μ infinitesimal). The solutions for this equation satisfy $\rho \approx \|(\alpha, \beta)\|\cos(\theta)$, so ρ and $\sqrt{1 + \rho^2}$ are finite. It follows that equation (1) holds with $\eta = \mu\sqrt{1 + \rho^2}$.

This ends the proof of 3.2.

REFERENCES

1. F. R. Dias Agudo, Lições de Análise Infinitesimal, Vol. 1, Cálculo Diferencial em $\mathbb{R}^n$, Livraria Escolar "Editora," 1969.
2. J. M. Henle, Tangent planes with infinitesimals, this MONTHLY, 91 (Aug.–Sept., 1984) 433–435.
3. H. J. Keisler, Foundations of Infinitesimal Calculus, Prindle, Weber & Schmidt, Boston, 1976.
4. K. D. Stroyan (in collaboration with W. A. J. Luxemburg), Introduction to the Theory of Infinitesimals, Academic Press, 1976.

Anthropomorphic Polygons

GODFRIED TOUSSAINT
School of Computer Science, McGill University, Montreal, Canada

We are concerned with a very special type of polygon in the Euclidean plane E^2 referred to as a *simple* (also *Jordan*) polygon. For any integer $n \geq 3$, we define a *polygon* or *n-gon* in the Euclidean plane E^2 as the figure $P = [x_1, x_2, \ldots, x_n]$ formed by n points $x_1, x_2, \ldots, x_n$ in E^2 and n line segments $[x_i, x_{i+1}]$, $i = 1, 2, \ldots, n-1$, and $[x_n, x_1]$. The points x_i are called the *vertices* of the *polygon* and the line segments are termed its *edges*.

DEFINITION. *A polygon P is called a simple polygon provided the only points of the plane that belong to two edges of P are the vertices of P. A simple polygon has a well defined interior and exterior. We will follow the convention of including the interior of a polygon when referring to P.*

DEFINITION (Meisters [**Me2**]). *A vertex x_i of P is called a principal vertex provided the diagonal $[x_{i-1}, x_{i+1}]$ intersects the boundary of P only at x_{i-1} and x_{i+1}.*

DEFINITION (Meisters [**Me1**]). *A principal vertex x_i of a simple polygon P is called an ear if the diagonal $[x_{i-1}, x_{i+1}]$ that bridges x_i lies entirely in P. We say that two ears x_i and x_j are nonoverlapping if* $\text{int}[x_{i-1}, x_i, x_{i+1}] \cap \text{int}[x_{j-1}, x_j, x_{j+1}] = \varnothing$.

The following *Two-Ears* Theorem was recently proved by Meisters [**Me1**].

THEOREM 1 (the *Two-Ears* Theorem, Meisters [**Me1**]). *Except for triangles every simple polygon P has at least two nonoverlapping ears.*

Meisters' proof by induction is both elegant and concise. However, given that a simple polygon can always be triangulated allows a one-sentence proof [**O'R**]. *Leaves* in the *dual-tree* of the triangulated polygon correspond to *ears* and every tree of two or more nodes must have at least two *leaves*.

This theorem is quite applicable in many situations. For example it establishes that there exists a stepwise procedure for "shrinking" a polygon P down to a triangle by deleting a vertex at each step, say x_i, and inserting $[x_{i-1}, x_{i+1}]$ in the place of $[x_{i-1}, x_i, x_{i+1}]$ while ensuring that at each step we retain a simple polygon P'. In actual fact, of course, we need only a "*one ear*" theorem to carry out such a procedure. The method is evident: locate an ear in P and "cut it off," then locate an ear in the remaining polygon of one less vertex and cut it off, and continue this process until the remaining polygon is a triangle. It is obvious that such a procedure could also be used as an algorithm for computing a triangulation of P. However care must be taken in converting this idea into an efficient algorithm. A straightforward approach of implementing this notion can result in a very slow algorithm. To determine if a vertex is or is not an ear may take $O(n)$ steps and we may have to visit $O(n)$ vertices to find and cut off an ear. Therefore, using a "brute force" approach we may have to perform $O(n^2)$ steps to cut off an ear and $O(n^3)$ steps to completely triangulate P in this manner. On the other hand, algorithms exist for triangulating simple polygons in time $O(n \log n)$ [**GJPT**] and $O(n \log\log n)$ [**TV**]. Once a triangulation is obtained the dual-tree can be determined in $O(n)$ time. Finally, an $O(n)$-time tree-traversal can prune off one *leaf* from the dual-tree at each step resulting in the cutting off of one *ear* from P at each step. It remains one of the most outstanding problems in computational geometry to determine if an $O(n)$ time algorithm exists for triangulating arbitrary simple polygons.

One question that arises is whether the "inverse" of the previous procedure is possible, i.e., does there always exist a stepwise procedure for "inflating" a simple polygon P until it is "as fat-as-possible" by deleting vertices from P one-at-a-time so that at each step we retain a simple polygon? We answer this question in the affirmative by proving that every non-convex polygon contains at least one *mouth*, but first we must define *mouth* and make more precise what we mean by "as-fat-as-possible."

DEFINITION. *A principal vertex x_i of a simple polygon P is called a mouth if the diagonal $[x_{i-1}, x_{i+1}]$ is an external diagonal, i.e., the interior of $[x_{i-1}, x_{i+1}]$ lies in the exterior of P.*

The convex hull of a simple polygon P will be denoted by $CH(P)$. The boundary (bd) of $CH(P)$ is a convex polygon. We now have a precise definition of "as-fat-as-possible," i.e., P is inflated until it becomes the convex hull of P.

THEOREM 2 (the *One-Mouth* Theorem). *Except for convex polygons every simple polygon P has at least one mouth.*

Proof. Construct the convex hull $CH(P)$. Since P is non-convex there must exist edges on $bd(CH(P))$ that are not edges of P. Each such edge forms the "lid" of a "pocket" of $CH(P)$. (Refer to FIG. 1.) We shall prove that in fact every such pocket yields a *mouth*. Let K_{ij} denote the pocket of $CH(P)$ determined by vertices x_i and x_j of P. Clearly $K_{ij} = [x_i, x_{i+1}, \ldots, x_j] \cup [x_j, x_i]$ forms itself a

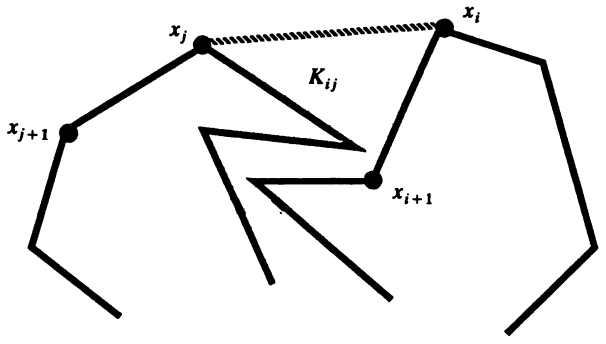


FIG. 1. Illustrating the proof of the *One-Mouth* Theorem.

simple polygon. By the *Two-Ears* Theorem K_{ij} must have two ears and since they are non-overlapping they cannot both occur at x_i and x_j. Therefore, at least one ear must occur at x_k for $i < k < j$. Obviously such an *ear* for K_{ij} is a *mouth* for P. Q.E.D.

While the above stepwise procedure for "inflating" a polygon P by "gobbling-up" mouths provides an algorithm for computing the *convex hull* of P this is not the best way to tackle this problem. Triangulating P does not appear to help here and a straightforward approach to "gobbling-up" mouths leads to an $O(n^3)$ time algorithm. On the other hand several $O(n)$ time algorithms for computing the convex hull of a simple polygon are known [**MA**], [**GY**], [**To**].

It is possible for a polygon to have many *ears* and only one *mouth* (FIG. 2(a)) and also many *mouths* and only one *ear* (FIG. 2(b)). Note that care is needed when speaking of *mouths* and *ears* as well *exposed* vertices, i.e., vertices of P that are also vertices of $CH(P)$. For example, Guggenheimer [**Gu**] states that a simple polygon has two *principal* vertices that are *exposed*. This is false and a counterexample due to Meisters [**Me2**] is illustrated in FIG. 2(c). This figure also illustrates that polygons exist which have precisely *one mouth* and *two ears*. In fact, these notions suggest some interesting families of simple polygons. Recall that no $O(n)$ time algorithm exists for triangulating an arbitrary simple polygon. However

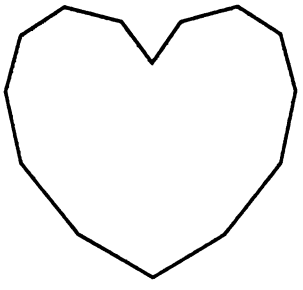


FIG. 2. (a) A polygon with only one *mouth* and many *ears*.

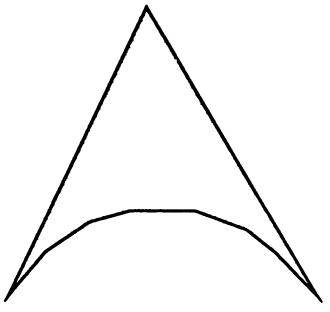


FIG. 2. (b) A polygon with only two *ears* and many *mouths*.

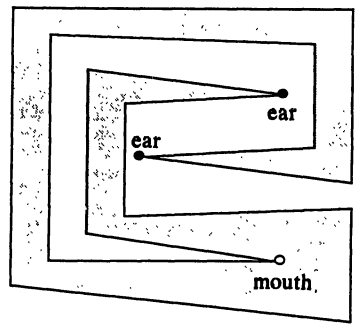


FIG. 2. (c) This polygon has precisely three *principal* vertices: two *ears* and one *mouth* and yet none of them are *exposed*.

certain special classes of simple polygons such as *star-shaped* ones do admit $O(n)$ time triangulation [**To2**]. We now define another such class of polygons.

DEFINITION. *A simple polygon P is called a one-mouth polygon provided it contains precisely one mouth.*

DEFINITION. *A simple polygon P is called a two-ear polygon provided it contains no more than two ears.*

DEFINITION. *A simple polygon P is called anthropomorphic provided it contains precisely two ears and one mouth.* (*See* FIG. 2(*c*).)

These three classes of polygons exhibit a good deal of structure as exemplified by the following theorem.

THEOREM 3. *The dual-tree of every triangulation of a two-ear polygon is a chain.*

Proof. Assume the polygon P has two ears and that the dual tree of some triangulation of P is not a chain. Then the tree must contain at least three leaves which is a contradiction. Q.E.D.

Theorem 3 allows us to triangulate the interior of a *two-ear* polygon of n vertices in $O(n)$ time as follows. Consider any vertex x_i of P. It is an easy matter to find another vertex x_j such that $[x_j, x_i]$ is an internal diameter of P in $O(n)$ time if indeed such a diagonal exists [**Le**]. Furthermore if such a diagonal does not exist then the diagonal $[x_{i-1}, x_{i+1}]$ is guaranteed to exist [**Le**]. In either case this diagonal partitions the polygon P into two polygons P_1 and P_2 each of which can be triangulated in $O(n)$ time starting at either $[x_j, x_i]$ or $[x_{i-1}, x_{i+1}]$. It suffices to realize that each diagonal can be inserted with a constant number of local angle tests.

A similar procedure can be used to triangulate the exterior of a *one-mouth* polygon. First we can use an $O(n)$ time algorithm for finding the convex hull of P [**To1**]. This will identify the two vertices x_i and x_j that form the "lid" of the pocket K_{ij} of $CH(P)$. One of the two ears of K_{ij} must occur at either x_i or x_j and can then be identified in a constant number of steps (i.e., independent of n). Triangulation of K_{ij} can then proceed as in the case of the *two-ear* polygon.

We have therefore established the following theorems.

THEOREM 4. *A one-mouth polygon can be externally triangulated in $O(n)$ time.*

THEOREM 5. *A two-ear polygon can be internally triangulated in $O(n)$ time.*

THEOREM 6. *An anthropomorphic polygon can be completely triangulated in $O(n)$ time.*

One additional computational problem that is of interest here concerns the *recognition* of these types of polygons. For example, whether a simple polygon is star-shaped or not can be determined in $O(n)$ time [**LP**]. By testing every vertex of a simple polygon to determine whether it is an ear or a mouth we can recognize *anthropomorphic* polygons in $O(n^2)$ time. However, using a more clever procedure we can reduce this complexity to $O(n)$ [**ST**].

Note added in proof: Since the writing of this note Bernard Chazelle at Princeton University has shown that arbitrary simple polygons can indeed be triangulated in linear time.

REFERENCES

[GJPT]. M. R. Garey, D. S. Johnson, F. P. Preparata, and R. E. Tarjan, Triangulating a simple polygon, *Information Processing Letters*, 7 (1978) 175–179.

[GY]. R. L. Graham, and F. F. Yao, Finding the convex hull of a simple polygon, *Journal of Algorithms*, 4 (1983) 324–331.

[Gu]. H. Guggenheimer, The Jordan and Schoenflies theorems in axiomatic geometry, this MONTHLY, 85 (1978) 753–756.

[LP]. D. T. Lee, and F. P. Preparata, An optimal algorithm for finding the kernel of a polygon, *Journal of the ACM*, 26 (1979) 415–421.

[Le]. L. S. Levy, Geometry: Modern Mathematics via the Euclidean Plane, Prindle, Weber, & Schmidt, Inc., Boston, 1970, pp. 141–142.

[MA]. D. McCallum and D. Avis, A linear time algorithm for finding the convex hull of a simple polygon, *Information Processing Letters*, 8 (1979) 201–205.

[Me1]. G. H. Meisters, Polygons have ears, this MONTHLY (June/July, 1975) 648–651.

[Me2]. _____, Principal vertices, exposed points, and ears, this MONTHLY, 87 (April 1980) 284–285.

[O'R]. J. O'Rourke, Art Gallery Theorems and Algorithms, Oxford University Press, 1987.

[ST]. T. Shermer and G. T. Toussaint, Anthropomorphic polygons can be recognized in linear time, Computing and Information, R. Janicki and W. W. Koczkodaj (Editors), Elsevier, 1989, pp. 117–123.

[TV]. R. E. Tarjan and C. J. Van Wyk, An $O(n \log\log n)$-time algorithm for triangulating simple polygons, *SIAM Journal on Computing*, 1988.

[To1]. G. T. Toussaint, A historical note on convex hull finding algorithms, *Pattern Recognition Letters*, 3 (January 1985) 21–28.

[To2]. G. T. Toussaint and D. Avis, On a convex hull algorithm for polygons and its application to triangulation problems, *Pattern Recognition*, 15 (1982) 23–29.

Inequalities for Areas Associated with Conics

W. A. DAY
Hertford College, Oxford

Let $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ be three points on the parabola $y^2 = 4ax$. The area of the triangle formed by the points is

$$\begin{aligned} A &= \frac{1}{2}|x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)| \\ &= \frac{1}{8a}|y_1^2(y_2 - y_3) + y_2^2(y_3 - y_1) + y_3^2(y_1 - y_2)| \\ &= \frac{1}{8a}|(y_1 - y_2)(y_2 - y_3)(y_3 - y_1)|. \end{aligned}$$

The tangents to the parabola at the three points are the lines

$$2a(x + x_1) = y_1 y, \qquad 2a(x + x_2) = y_2 y, \qquad 2a(x + x_3) = y_3 y,$$

and these intersect at the points

$$\left(\frac{y_1 y_2}{4a}, \frac{y_1 + y_2}{2}\right), \left(\frac{y_2 y_3}{4a}, \frac{y_2 + y_3}{2}\right), \left(\frac{y_3 y_1}{4a}, \frac{y_3 + y_1}{2}\right).$$

Thus, the area of the triangle formed by the tangents is

$$\begin{aligned} A' &= \frac{1}{16a}|y_1 y_2(y_1 - y_2) + y_2 y_3(y_2 - y_3) + y_3 y_1(y_3 - y_1)| \\ &= \frac{1}{16a}|(y_1 - y_2)(y_2 - y_3)(y_3 - y_1)| = \tfrac{1}{2}A. \end{aligned}$$

In summary: *the area of the triangle formed by three tangents to a parabola is half that of the triangle formed by joining their points of contact.*

This conclusion has long been known, and is to be found, for example, in Salmon's *Conic Sections* [1]. Having set the proof as an undergraduate exercise, I was led to ask about the relationships between A' and A for the ellipse and the hyperbola. My purpose is to point out that *the area of the triangle formed by three tangents to an ellipse is strictly greater than half that of the triangle formed by joining their points of contact*, and that *the area of the triangle formed by three tangents to a hyperbola is strictly less than half that of the triangle formed by joining their points of contact, provided that all three points of contact lie on the same branch of the hyperbola.* (It is easily seen that the latter conclusion is generally false if the points of contact do not lie on the same branch).

I shall establish these conclusions by proving somewhat more, namely, that *for the ellipse*

$$A' \geq \frac{A}{2\left(1 - \left(\frac{A}{2ab}\right)^{2/3}\right)^{3/2}}, \tag{1}$$

while for the hyperbola

$$A' \leq \frac{A}{2\left(1 + \left(\frac{A}{2ab}\right)^{2/3}\right)^{3/2}}, \tag{2}$$

where a and b are the lengths of the semi-major and semi-minor axes.

The proof depends on two instances of an elementary inequality, which says that *if* p, q, r, u, v, w *are any nonnegative numbers then*

$$(pqr)^{1/3} + (uvw)^{1/3} \leq (p+u)^{1/3}(q+v)^{1/3}(r+w)^{1/3}. \tag{3}$$

To verify (3) we observe that

$$\left((pqr)^{1/3} + (uvw)^{1/3}\right)^3 = pqr + 3\left(p^2q^2r^2uvw\right)^{1/3} + 3\left(pqru^2v^2w^2\right)^{1/3} + uvw.$$

However, the arithmetic-geometric mean inequality implies the inequalities

$$3\left(p^2q^2r^2uvw\right)^{1/3} \leq pqw + qru + rpv,$$

$$3\left(pqru^2v^2w^2\right)^{1/3} \leq pvw + qwu + ruv.$$

Hence,

$$\begin{aligned}\left((pqr)^{1/3} + (uvw)^{1/3}\right)^3 &\leq pqr + pqw + qru + rpv + pvw + qwu + ruv + uvw \\ &= (p+u)(q+v)(r+w),\end{aligned}$$

and (3) follows on taking the cube root of each side.

To obtain the first of the required special cases, let α, β, γ be any real numbers and let

$$\begin{aligned} p &= \cos^2\alpha, & q &= \cos^2\beta, & r &= \cos^2\gamma, \\ u &= \sin^2\alpha, & v &= \sin^2\beta, & w &= \sin^2\gamma. \end{aligned}$$

Then (3) tells us that

$$(\cos^2\alpha\cos^2\beta\cos^2\gamma)^{1/3} + (\sin^2\alpha\sin^2\beta\sin^2\gamma)^{1/3} \leq 1. \tag{4}$$

On the other hand, if we take

$$\begin{aligned} p = q = r = 1, \\ u = \sinh^2\alpha, \qquad v = \sinh^2\beta, \qquad w = \sinh^2\gamma, \end{aligned}$$

we find the inequality

$$1 + (\sinh^2\alpha\sinh^2\beta\sinh^2\gamma)^{1/3} \leq (\cosh^2\alpha\cosh^2\beta\cosh^2\gamma)^{1/3}, \tag{5}$$

which is the second of the special cases.

We are now in position to verify (1). To this end we adopt the parametric description

$$x = a\cos t, \qquad y = b\sin t, \qquad (t \in [0, 2\pi))$$

of the ellipse. The area of the triangle formed by the points corresponding to the

values t_1, t_2, t_3 of the parameter t proves to be

$$\begin{aligned} A &= \frac{1}{2}ab\left|\sin(t_1 - t_2) + \sin(t_2 - t_3) + \sin(t_3 - t_1)\right| \\ &= ab\left|\sin\frac{1}{2}(t_1 - t_2)\cos\frac{1}{2}(t_1 - t_2)\right. \\ &\qquad \left. -\sin\frac{1}{2}(t_1 - t_2)\cos\frac{1}{2}(t_1 + t_2 - 2t_3)\right| \\ &= 2ab\left|\sin\frac{1}{2}(t_1 - t_2)\sin\frac{1}{2}(t_2 - t_3)\sin\frac{1}{2}(t_3 - t_1)\right|. \end{aligned} \tag{6}$$

The tangents to the ellipse at the points corresponding to the values t_1, t_2, t_3 are the lines

$$\frac{x\cos t_1}{a} + \frac{y\sin t_1}{b} = 1$$

and so forth, and the area of the triangle formed by these lines proves to be

$$\begin{aligned} A' &= \frac{ab\left|\sin(t_1 - t_2) + \sin(t_2 - t_3) + \sin(t_3 - t_1)\right|^2}{2\left|\sin(t_1 - t_2)\sin(t_2 - t_3)\sin(t_3 - t_1)\right|} \\ &= \frac{8ab\left|\sin\frac{1}{2}(t_1 - t_2)\sin\frac{1}{2}(t_2 - t_3)\sin\frac{1}{2}(t_3 - t_1)\right|^2}{\left|\sin(t_1 - t_2)\sin(t_2 - t_3)\sin(t_3 - t_1)\right|} \\ &= ab\left|\tan\frac{1}{2}(t_1 - t_2)\tan\frac{1}{2}(t_2 - t_3)\tan\frac{1}{2}(t_3 - t_1)\right|, \end{aligned}$$

and, hence,

$$A' = \frac{1}{2}A\left|\sec\frac{1}{2}(t_1 - t_2)\sec\frac{1}{2}(t_2 - t_3)\sec\frac{1}{2}(t_3 - t_1)\right|. \tag{7}$$

On combining (6) and (7) with the inequality (4), we see that

$$\left(\frac{A}{2A'}\right)^{2/3} + \left(\frac{A}{2ab}\right)^{2/3} \leq 1,$$

and this is equivalent to (1).

To verify (2) we adopt the parametric description

$$x = a\cosh t, \qquad y = b\sinh t, \qquad (t \in \mathbb{R})$$

of a branch of the hyperbola. In this case, the area of the triangle formed by the points corresponding to the values t_1, t_2, t_3 of the parameter t turns out to be

$$A = 2ab\left|\sinh\frac{1}{2}(t_1 - t_2)\sinh\frac{1}{2}(t_2 - t_3)\sinh\frac{1}{2}(t_3 - t_1)\right|,$$

while the counterpart to (7) turns out to be

$$A' = \frac{1}{2}A\left|\operatorname{sech}\frac{1}{2}(t_1 - t_2)\operatorname{sech}\frac{1}{2}(t_2 - t_3)\operatorname{sech}\frac{1}{2}(t_3 - t_1)\right|.$$

On appealing to (5), we obtain the inequality

$$1 + \left(\frac{A}{2ab}\right)^{2/3} \leq \left(\frac{A}{2A'}\right)^{2/3},$$

which is equivalent to (2).

REFERENCE

1. G. Salmon, A Treatise on Conic Sections, 6th edition, Longman, Green, and Co., London, 1879, p. 212.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

The Derivative á la Carathéodory

STEPHEN KUHN
Department of Mathematics, University of Tennessee, Chattanooga, TN 37403

1. Introduction. Augustin-Louis Cauchy would be pleased. Each year we introduce our elementary analysis students to the notion of the derivative essentially as he gave it to us in 1823 (for details see, e.g. [**4**, **7**]). But there is another, less well known, characterization of the derivative which appears in the last textbook [**2**] written by Constantin Carathéodory (1873–1950). This formulation is not only elegant but useful on both theoretical and pedagogical grounds. The proofs of many important theorems concerning differentiability become significantly easier and, in the process, some of the less enlightening details are rightly submerged.

Carathéodory's formulation also gives us a much clearer view of the fact that continuity is essential for differentiability; indeed, the definition itself contains the necessary continuity. This formulation, which shifts the details from the theory of limits to the theory of continuous functions, requires that our students develop a clearer understanding of continuity than is typical and it demands that we reevaluate this understanding and continually reinforce it.

We believe that Carathéodory's insight deserves to be better known and we hope that the present article will help in that effort. We also believe, however, that since it is somewhat more subtle and less practical from a computational point of view than is the standard definition, Carathéodory's approach should be neither the first nor the only one that students see. But we also believe that all elementary analysis students can appreciate its power and should be exposed to it; in fact even the more perceptive first-year calculus students have something to gain from working with it.

2. Carathéodory's derivative formulation. After the usual definitions and theorems about limits and continuity are presented in a standard elementary real analysis course, the definition of the derivative, essentially as given to us by Cauchy, is given for functions of a single variable. Typically it is presented in both the following forms:

Definition. The value of the derivative of the function f at a is the number

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

if the limit exists; equivalently,

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

if the limit exists.

Carathéodory realized the importance of the fact that the criterion for continuity is apparent in the second definition above and he proposed the following definition for the derivative at a point; see [**2**, p. 119] or [**8**, p. 58].

CARATHÉODORY'S DEFINITION OF THE DERIVATIVE. The function f, defined on the open interval U, is said to be differentiable at the point $a \in u$ if there exists a function ϕ_a that is continuous at $x = a$ and satisfies the relation $f(x) - f(a) = \phi_a(x)(x - a)$ for all $x \in U$.

We will usually write $\phi(x)$ instead of $\phi_a(x)$, since there seems to be little chance of confusion, but we must remember that the function ϕ depends on the point a. Geometrically, of course, when $x \neq a$, $\phi_a(x)$ is the slope of the secant line through the points $(x, f(x))$ and $(a, f(a))$. This alternative definition emphasizes the fact that the slopes of the secant lines, by way of which we initially arrive at the tangent line, approach the tangent line in a continuous manner; we rarely state this important fact explicitly, but we should.

Clearly all of the subtlety of the limit in the standard definition of the derivative is now crouching just below the surface, hiding in the subtlety of the continuity of the function ϕ.

This definition can be used as is for complex-valued functions of a complex variable; in fact, it is in precisely this context that Carathéodory introduced it in [2]. It is interesting to note, however, that this formulation does not appear in Carathéodory's 1918 text [3] on real-valued functions.

As pointed out in [2] and [8], two immediate consequences of this formulation are:

- If f has a derivative at a, then f is continuous at a.
- There is at most one function ϕ satisfying the definition; so if $f'(a)$ exists, $f'(a) = \phi(a)$.

These two corollaries of the definition are both expected and their proofs are trivial, assuming only elementary results about continuous functions. The power of Carathéodory's approach does not become apparent until we apply it to prove some deeper theorems. In the following sections we offer proofs of several standard, basic theorems which we hope will demonstrate the value of this formulation.

3. An indicative example: the chain rule. Shortly after the introduction of his definition of the derivative, Cauchy used it to prove the chain rule (though his proof was incomplete). It seems appropriate then that we do the same with Carathéodory's definition. "How incisive Carathéodory's formulation is may be seen by applying it to the proof of the chain rule for the derivative of a composite function" [8, p. 59]. Many of our students will appreciate the pithy elegance of this proof, which is slightly revised from [8].

THEOREM 1. CHAIN RULE. *If f is differentiable at the point a and g is differentiable at the point $b = f(a)$, then $h = g \circ f$ is differentiable at a and $h'(a) = g'(f(a))f'(a)$.*

Proof. Since f is differentiable at a there is a function ϕ, continuous at a and defined in an open interval V containing a, with $f(x) - f(a) = f(x) - b = \phi(x)(x - a)$ for all $x \in V$. Similarly we get a function ψ, continuous at b and defined in an open interval U containing a, for which $g(x) - g(b) = \psi(x)(x - b)$ for all $x \in U$.

Then

$$h(x) - h(a) = g(f(x)) - g(f(a)) = \psi[f(x)][f(x) - b]$$
$$= [\psi \circ f(x)]\phi(x)(x - a),$$

for all $x \in V$ with $f(x) \in U$. Since $(\psi \circ f)\phi$ is continuous at a and has the value $g'(f(a))f'(a)$ there, the proof is complete. □

4. Additional examples. Carathéodory's characterization also leads to some sharp, concise proofs of other important theorems and we hope that the examples below will demonstrate the merit of Carathéodory's approach.

THEOREM 2. INVERSE FUNCTION THEOREM. *Let f be continuous and strictly monotonic on an open interval I containing c and assume that $f'(c)$ exists and is nonzero. Then $g = f^{-1}$ is differentiable at $d = f(c)$ and $g'(d) = [f'(c)]^{-1}$.*

Proof. (Our proof here is from [**8**, p. 61].) We have $f(x) - f(c) = \phi(x)(x - d)$ for all x in I with ϕ continuous and $\phi(x) \neq 0$ for all x in I. Let V be the open interval which is the domain of g. Then $y - d = f(g(y)) - d = f(g(y)) - f(c) = \phi[g(y)][g(y) - c]$ for all y in V; thus $[g(y) - c] = (1/\phi[g(y)])(y - d)$ for all y in V. Since g is continuous on V, $1/(\phi \circ g)$ is continuous there as well, and the theorem follows. □

If Carathéodory's alternative were powerful enough to prove theorems such as the two above, but too powerful or cumbersome to handle the earliest calculational results of the subject, we would be less likely to introduce it. In fact, it also manages these earlier theorems with ease and even some grace.

THEOREM 3. LINEARITY AND THE PRODUCT AND POWER RULES. *If f and g are differentiable at a, k is a constant, and n is a natural number, then*:

i) $[kf + g]'(a) = kf'(a) + g'(a)$.

ii) $(fg)'(a) = f(a)g'(a) + g(a)f'(a)$.

iii) *the derivative of* $f(x) = x^n$ *at* a *is* $f'(a) = na^{n-1}$.

Proof. We assume that there exists ϕ, ψ, continuous at a and defined on open intervals U and V, respectively, containing a, such that $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in U$ and $g(x) - g(a) = \psi(x)(x - a)$ for all $x \in V$.

i) Then

$$[kf + g](x) - [kf + g](a) = k[f(x) - f(a)] + [g(x) - g(a)]$$
$$= k\phi(x)(x - a) + \psi(x)(x - a)$$
$$= [k\phi(x) + \psi(x)](x - a), \quad \text{for all } x \in U \cap V.$$

ii) For all $x \in U \cap V$,

$$(fg)(x) - (fg)(a) = f(x)g(x) - f(a)g(a)$$
$$= f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)$$
$$= f(x)[g(x) - g(a)] + [f(x) - f(a)]g(a)$$
$$= f(x)\psi(x)(x - a) + \phi(x)(x - a)g(a)$$
$$= [f(x)\psi(x) + \phi(x)g(a)](x - a).$$

iii) Since $x^n - a^n = [x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}](x - a)$ for all x, we have $\phi(x) = x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}$ and hence $\phi(a) = na^{n-1}$. □

Finally, we easily derive another basic but very useful result.

THEOREM 4. CRITICAL POINT THEOREM. *Let f be defined on an open interval I containing a. If $f(a)$ is an extreme value then a is a critical point (i.e., $f'(a) = 0$ or $f'(a)$ does not exist).*

Proof. We only prove the theorem for $f(a)$ a maximum and we do this by assuming $f'(a)$ exists and proving that $f'(a) = 0$.

For some open subinterval U of I we have a function ϕ, continuous at a, such that $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in U$. Let $c \in U$, $c < a$. Then $f(c) - f(a) = \phi(c)(c - a)$, and since $f(a)$ is a maximum in U, $f(c) \leq f(a)$. So $f(c) - f(a) \leq 0$ and $c - a < 0$; hence $\phi(c) \geq 0$. Similarly, for $d \in U$, $d > a$, we have $f(d) - f(a) \leq 0$ and so $\phi(d) \leq 0$.

Since $\phi(c) \geq 0$ for all $c \in U$, $c < a$, and $\phi(d) \leq 0$ for all $d \in U$, $d > a$, the continuity of ϕ at a forces us to conclude that $\phi(a) = 0$. □

It is surprising that so few elementary analysis texts even mention this alternate view of the derivative. Of ten undergraduate/graduate analysis texts we examined only Rosenlicht [**12**], makes use of it, although he does not mention Carathéodory by name; he imbeds it in his proof of the chain rule for one variable and brings it out in the open for partial differentiation.

There are several analysis texts, including Protter and Morrey [**10**], which offer $f'(a) + r(x; a)$, where $r(x; a) \to 0$ as $x \to a$, instead of $\phi_a(x)$, and use it to good advantage in the proofs of several theorems. This approach has the intended additional benefit of making transparent the linear approximation of the tangent line.

The simplification of additional proofs using this definition and the extension of it to functions of several variables would make interesting projects for enterprising analysis students.

5. Carathéodory's approach in first-year calculus? It is clear that Carathéodory's formulation demands too much of most first-year calculus students to make its presentation worthwhile. Nevertheless, it can be of value, for example, for students in honors calculus courses. The chain rule will serve as a good representative example of the relative improvements in such courses afforded by the proofs given here.

Many current standard calculus texts ([**11**] is typical in this regard) relegate the more delicate details of the proof to an appendix, which will be read by few students. At the extreme, [**5**] puts the entire proof in the appendix. At the other extreme, [**9**] gives a $1\frac{1}{4}$-page "simpler proof that is valid for many functions" and then another $1\frac{3}{4}$-page complete proof (both proofs are within the same section of the text!). Neither of these routes is necessary, however. Using Carathéodory's definition we get a very clean, understandable proof of the chain rule which can be presented to perceptive first-year students in its entirety without interrupting the flow of the text.

Of 18 standard calculus texts we examined, only one, Gillett [**6**], refers to Carathéodory's formulation of the derivative at all (exercise 66, p. 127), but no

mention is made of Carathéodory himself. There are a few calculus texts, including Boyce and DiPrima [1], which present the variation mentioned at the end of the previous section.

(Note: Young [13] contains interesting biographical information on Carathéodory, from the viewpoint of one of his students and strong admirers.)

Acknowledgement. Thanks to the referee for several helpful suggestions.

REFERENCES

1. William Boyce and Richard DiPrima, Calculus, Wiley, New York, 1988.
2. Constantin Carathéodory, Theory of Functions of a Complex Variable, vol. 1, Chelsea, New York, 1954.
3. _____, Vorlesungen über Reele Functionen, Chelsea Publishing Company, New York, 1968.
4. C. H. Edwards, Jr., The Historical Development of the Calculus, Springer-Verlag, New York, 1979.
5. Robert Ellis and Denny Gulick, Calculus with Analytic Geometry, Harcourt Brace Jovanovich, New York, 1978.
6. Philip Gillett, Calculus and Analytic Geometry, third ed., D. C. Heath, Lexington, Mass., 1988.
7. Judith V. Grabiner, The Origin of Cauchy's Rigorous Calculus, MIT Press, Cambridge, Mass., 1981.
8. Maurice Heins, Complex Function Theory, Academic Press, New York, 1968.
9. Louis Leithold, The Calculus with Analytic Geometry, fifth ed., Harper and Row, New York, 1986.
10. M. H. Protter and C. B. Morrey, A First Course in Real Analysis, Springer-Verlag, New York, 1977.
11. Edwin Purcell and Dale Varburg, Calculus with Analytic Geometry, fifth ed., Prentice-Hall, Englewood Cliffs, N.J., 1987.
12. Maxwell Rosenlicht, Introduction to Analysis, Scott Foresman, Glenview, Ill., 1968.
13. Laurence Young, Mathematicians and Their Times, North-Holland, Amsterdam, 1981.

Positive Definite Matrices and Sylvester's Criterion

GEORGE T. GILBERT
Department of Mathematics, Texas Christian University, Fort Worth, TX 76129

Sylvester's criterion states that a symmetric (more generally, Hermitian) matrix is positive definite if and only if its principal minors are all positive. The theorem is especially useful when employing the Second Derivative Test for local extrema of functions of several variables. When the proof of the sufficiency of positive principal minors is included in a linear algebra text, it is usually based on Gaussian elimination [4, pp. 331–332] or on the reduction theory for quadratic forms [2, pp. 328–329]. In either case, the heart of such proofs is a matrix computation. This note presents a proof which, while still very dependent on matrices, makes more use of ideas from the theory of vector spaces. In a one semester course in linear algebra, one often has the feeling of developing the machinery of abstract vector spaces without ever really making use of it. The proof here is short, requires only concepts that are standard in beginning linear algebra, and provides an opportu-

nity for the student to see the abstract theory applied to a concrete problem. A less direct proof in a similar spirit is given in [3].

We begin by stating the prerequisite theorems and definitions. The first theorem we shall need is the following:

THEOREM 1. *Let W_1 and W_2 be finite-dimensional subspaces of a vector space V. Then*

$$\dim(W_1 + W_2) + \dim(W_1 \cap W_2) = \dim W_1 + \dim W_2.$$

Proof. One simply extends a basis for $W_1 \cap W_2$ to a basis for W_1 and to a basis for W_2. The bases combine to form a basis for $W_1 + W_2$. See [2, p. 46] or [4, pp. 199–200] for details.

Since the determinants of similar matrices are equal, we also have

PROPOSITION. *The determinant of a diagonalizable matrix is the product of its eigenvalues.*

The next theorem provides one of many opportunities for nontrivial inductive proofs in a linear algebra course.

THEOREM 2 (Spectral Theorem). *Let A be an $n \times n$ real, symmetric matrix. Then all eigenvalues of A are real and there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

Proof. See [2, pp. 312–314] or [4, pp. 295–296] for a proof.

Having stated the background we require, we proceed to positive definite matrices.

DEFINITION. A real, symmetric matrix is *positive definite* if for every nonzero vector in $\mathbb{R}^n$, $\mathbf{v}^t A\mathbf{v} > 0$.

The proof of the next theorem is a routine exercise.

THEOREM 3. *A real, symmetric matrix A is positive definite if and only if all its eigenvalues are positive.*

Once we define principal minors, we turn to the proof of Sylvester's criterion.

DEFINITION. Let A be an $n \times n$ matrix. For $1 \leq k \leq n$, the *kth principal submatrix* of A is the $k \times k$ submatrix formed from the first k rows and first k columns of A. Its determinant is the *kth principal minor*.

THEOREM (Sylvester's Criterion). *A real, symmetric matrix is positive definite if and only if all its principal minors are positive.*

It is easy to prove that positive principal minors are necessary by showing that each kth principal submatrix is positive definite, hence has positive determinant by Theorem 3 and the Proposition. Our proof that positive principal minors implies a matrix is positive definite proceeds through two lemmas, instructive in their own rights. At each stage we use vector space methods, although, of course, parallel matrix proofs are possible.

LEMMA 1. *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *be a basis of a vector space* V. *Suppose* W *is a* k*-dimensional subspace of* V. *If* $m < k$, *there exists a nonzero vector in* W *which is a linear combination of* $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$.

Proof. Since $\dim W + (n - m) = k + n - m > n$, W has nontrivial intersection with the span of $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ by Theorem 1.

The difference between our proof and others lies mainly in the next lemma, which is nearly a special case of the Courant-Fischer "min-max theorem" [3, p. 179].

LEMMA 2. *Let* A *be an* $n \times n$ *real, symmetric matric. If* $\mathbf{w}^t A\mathbf{w} > 0$ *for all nonzero vectors* $\mathbf{w}$ *in a* k*-dimensional subspace* W *of* $\mathbb{R}^n$, *then* A *has at least* k *positive eigenvalues (counting multiplicity).*

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Suppose further that the first m eigenvalues are positive and the rest are not. If $m < k$, then Lemma 1 implies there exists a nonzero vector $\mathbf{w}$ in W which may be written

$$\mathbf{w} = c_{m+1}\mathbf{v}_{m+1} + \cdots + c_n\mathbf{v}_n.$$

We have

$$\mathbf{w}^t A\mathbf{w} = \sum_{j,k=m+1}^{n} c_j c_k \lambda_k \mathbf{v}_j^t \cdot \mathbf{v}_k = c_{m+1}^2 \lambda_{m+1} + \cdots + c_n^2 \lambda_n,$$

since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal set. But then $\mathbf{w}^t A\mathbf{w} \leq 0$, a contradiction. Thus $m \geq k$, as desired.

We complete the proof of sufficiency in Sylvester's criterion by induction. For $n = 1$, the result is trivial. Assume the sufficiency of positive principal minors for $(n - 1) \times (n - 1)$ real, symmetric matrices. If A is an $n \times n$ real, symmetric matrix with positive principal minors, then, by the inductive hypothesis, its $(n - 1)$st principal submatrix is positive definite. Let W be the $(n - 1)$-dimensional subspace of $\mathbb{R}^n$ consisting of those vectors whose last coordinate is zero. Then for any nonzero vector $\mathbf{w}$ in W, $\mathbf{w}^t A\mathbf{w} > 0$. Lemma 2 now implies that A has at least $n - 1$ positive eigenvalues (counting multiplicity). We apply the Proposition above and the fact that $\det A > 0$ to conclude that A has n positive eigenvalues. This completes the proof of Sylvester's criterion.

Acknowledgements. The author wishes to thank the referee for suggesting several improvements and for bringing [3] to his attention.

REFERENCES

1. N. V. Efimov, E. R. Rozendorn, Linear Algebra and Multidimensional Geometry, Mir Publishers, Moscow, 1975.
2. K. Hoffman and R. Kunze, Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1971.
3. R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
4. G. Strang, Linear Algebra and Its Applications, 3rd ed., Harcourt Brace Jovanovich, Inc., San Diego, 1988.

The Gibbs Phenomenon for Piecewise-Linear Approximation

J. Foster and F. B. Richards
Department of Mathematics, Weber State College, Ogden, UT 84408

In 1899 J. W. Gibbs [1] of Yale, in response to a letter in *Nature* by the American physicist A. Michelson [2], presented a result about Fourier series that now goes by the name of the Gibbs phenomenon. Michelson had complained about an undesired "overshoot" effect that occurred whenever he approximated a function having a jump discontinuity by a finite Fourier series. Gibbs was able to show that this overshoot does not disappear as the number of terms in the series becomes arbitrarily large. Rather, on each side of the discontinuity, it approaches a constant g times one-half the size of the jump, where

$$g = (2/\pi)\int_0^{\pi} \frac{\sin x}{x}\, dx - 1 = 0.17897974$$

independent of the function being approximated.

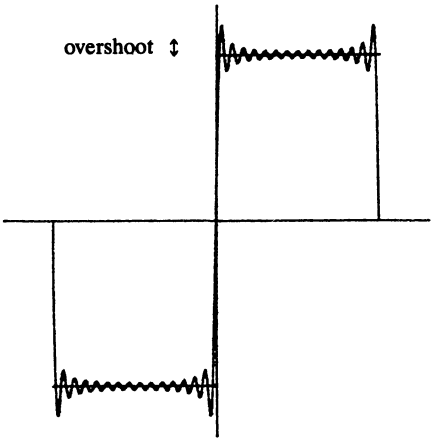


Fig. 1. Approximation of a square wave by a 30-term Fourier series.

This phenomenon is familiar to every student of differential equations or Fourier series, but textbooks often give the impression that it is a special quirk of Fourier series. The point we wish to make in this note is that, on the contrary, the Gibbs phenomenon has little to do with the use of trigonometric polynomials as approximating functions but is primarily a consequence of the fact that Fourier series are *best least squares approximations*; and that we should expect to see an effect of this nature whenever we employ least squares approximations to a jump-discontinuous function, even if the approximating functions are not trigonometric polynomials.

To illustrate, we will prove an analog of the Gibbs phenomenon when the approximating functions are piecewise-linear. The computations are straightforward and would make a good classroom exercise or project.

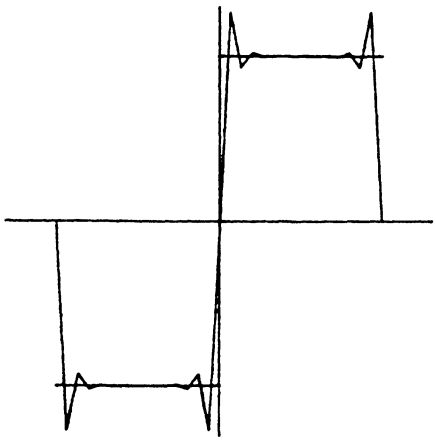


FIG. 2. Best least squares approximation of a square wave by a piecewise-linear function with 30 equally spaced nodes per period.

Define

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ -1, & -1 < x < 0 \end{cases}$$

and extend f by periodicity to the entire real line. The value of $f(x)$ when x is integral is immaterial. We will find the unique periodic continuous function $y^{(n)}$ which is linear on each interval $[k/n, (k+1)/n]$, n a fixed positive integer, and which best approximates f in the L_2 norm on $[-1, 1]$. The numbers k/n are called the *nodes* of y^n.

Under the notation $y_k = y^{(n)}(k/n)$ it should be evident from symmetry that $y_{-k} = -y_k$ and $y_{-n} = y_0 = y_n = 0$. So it will suffice to find numbers $y_0, y_1, \ldots, y_n$ that minimize

$$\int_0^1 \left(y^{(n)}(x) - 1\right)^2 dx.$$

It is easy to verify that for $k/n \leqslant x \leqslant (k+1)/n$,

$$y^{(n)}(x) = (nx - k)(y_{k+1} - y_k) + y_k.$$

If we let $w_k = y_k - 1$ we then have

$$y^{(n)}(x) - 1 = (nx - k)(w_{k+1} - w_k) + w_k$$

and

$$\int_{k/n}^{(k+1)/n} \left(y^{(n)}(x) - 1\right)^2 dx = \left(w_{k+1}^2 + w_{k+1}w_k + w_k^2\right)/3n.$$

Then

$$\int_0^1 \left(y^{(n)}(x) - 1\right)^2 dx = \left(2\sum_{k=1}^{n-1} w_k^2 + \sum_{k=1}^{n-2} w_{k+1}w_k + 2 - w_1 - w_{n-1}\right)/3n.$$

By setting the partial derivatives with respect to $w_1, \ldots, w_{n-1}$ equal to zero we

obtain the *normal equations*

$$4w_1 + w_2 = 1$$
$$w_{k-1} + 4w_k + w_{k+1} = 0 \quad (2 \leqslant k \leqslant n-2)$$
$$w_{n-2} + 4w_{n-1} = 1.$$

The general equation in this system is a second-order linear homogeneous difference equation whose general solution is $w_k = c_1 r_1^k + c_2 r_2^k$ where $r_1 = -2 + \sqrt{3}$ and $r_2 = -2 - \sqrt{3}$ are the roots of the characteristic equation $r^2 + 4r + 1 = 0$. The extreme equations enable us to solve for c_1 and c_2:

$$c_1 = (r_2^n - 1)/(r_1^n - r_2^n)$$
$$c_2 = (r_1^n - 1)/(r_2^n - r_1^n).$$

The solution of the system is then

$$w_k = -(r_1^k + r_1^{n-k})/(1 + r_1^n).$$

Since $0 < -r_1 < 1$, it is easy to see that the sequence

$$(-1)^{k+1} w_k = (-r_1)^k (r_1^{n-2k} + 1)/(1 + r_1^n)$$

is positive for $1 \leqslant k \leqslant n/2$. It is also decreasing, because

$$(-1)^{k+1} w_k - (-1)^{k+2} w_{k+1} = (-r_1)^k (r_1 + 1)(r_1^{n-2k-1} + 1)/(1 + r_1^n) > 0$$

for $1 \leqslant k < n/2$. Hence the maximum "overshoot" (that is, the deviation of $y^{(n)}(x)$ from $f(x)$) occurs when $k = 1$ and its value is $w_1 = w_1^{(n)} = -(r_1 + r_1^{n-1})/(1 + r_1^n)$. To see that this overshoot doesn't disappear as n gets large, we compute

$$\lim_{n \to \infty} w_1^{(n)} = -r_1 = 2 - \sqrt{3}.$$

If we define the *Gibbs constant* to be the overshoot divided by one-half the magnitude of the jump, then we have shown that the Gibbs constant for approximation by piecewise-linear functions is $2 - \sqrt{3} = 0.26794919$, somewhat larger than the classical Gibbs constant.

Note 1. We leave it to the reader to verify that the Gibbs constant is independent of the function f.

Note 2. Continuous piecewise-linear functions are referred to in approximation theory as *splines of degree* 1. The second author of this paper has shown that a Gibbs phenomenon holds for splines of higher degree [**3**].

REFERENCES

1. J. Willard Gibbs, letter to the editor, *Nature* 59 (April 27, 1899) 606. (This letter corrected an error he had made in a letter of December 29, 1898).
2. Albert A. Michelson, letter to the editor, *Nature* 58 (October 6, 1898) 544–45.
3. F. B. Richards, A Gibbs phenomenon for spline functions (submitted).

Teaching and Testing Mathematics Reading

CARL C. COWEN
Department of Mathematics, Purdue University, West Lafayette, IN 47907

When we discuss teaching undergraduate mathematics with our colleagues, we frequently distinguish two sorts of goals: first, we want students to be able to do computations related to the course and, second, we want students to know the theory from the course. In a calculus course, we want students to be able to integrate polynomials, so in class we tell them how, and on the exam, we give them polynomials to integrate. In linear algebra class, we want the students to know the theorems concerning determinants, so we prove the theorems from the book, and on the exam, we ask them to prove that if A is an invertible matrix, then $\det(A^{-1}) = 1/\det(A)$. Our hallway conversations elicit general agreement that most of our students do learn to do the calculations we put on the exam and most don't master the theory to the point that they can write a coherent proof of any but the most trivial theorems.

I want to argue in this note that there is another goal, between these in accessibility, that should be a conscious part of our teaching effort:

We should teach our students to read and understand mathematics.

And to reinforce our effort, we must test their ability to do so. Very few of our students will ever, after leaving our courses, have to give a formal proof of a theorem and few will have to do something so mundane as "Find the definite integral." On the other hand, many of them *will* have to read and understand mathematical writing to apply new ideas to the problems of their jobs. Moreover, learning to read and understand mathematics will make the further step of learning to prove theorems more possible and will have benefits for every course that follows. We need to convince our students that, in the long run, understanding what the theorems mean will be easier and have a bigger payoff than simply memorizing their statements. I do not believe that our current goals should be dropped; they are an important part of every mathematics course. On the other hand, I do not believe that learning to read can be left simply to chance; that's what we've been doing and it hasn't worked!

If you need evidence that we have a problem, let one of your "B" students explain the statement of a theorem from a homework exercise that they have not been able to prove or have them explain the statement and proof of a theorem from a section in the book that you have skipped. My students, at least, do not have the innate ability to read and understand what they have read. When I ask them to read a problem and explain it to me, the majority simply recite the same words back again.

To help them learn to read, we must give them opportunities to practice. In class, it is helpful to discuss the meaning of a new theorem before proving it. It is especially important to discuss hypotheses. Many students have not learned what role the hypothesis of a theorem plays. You might give several examples of objects that could satisfy the hypotheses of a theorem and get your class to tell you which actually do. Asking them to restate the conclusion in a particular instance is frequently a helpful way to deal with the conclusion. Drawing out corollaries that

have not yet been stated is another way to deal with the conclusion. For example, in an advanced linear algebra course, I prove the following theorem:

THEOREM. *If A is an n by n matrix and μ is a number such that*

$$\|A\| < |\mu|$$

then $\mu I - A$ is invertible.

Comment: This theorem is true for any submultiplicative norm on matrices; the norm used below is the one-norm:

$$\|A\| = \max\left\{ \sum_{i=1}^{n} |a_{ij}| : j = 1, \ldots, n \right\}$$

After writing the statement of the theorem on the blackboard, we can have the following (idealized) discussion:

Instructor: What is the one-norm of the matrix

$$A = \begin{pmatrix} .7 & -.2 \\ .5 & 1.4 \end{pmatrix}?$$

Student: $\|A\| = 1.6$

Instructor: Good! What does this theorem say about the matrix

$$2I - A = \begin{pmatrix} 1.3 & .2 \\ -.5 & .6 \end{pmatrix}?$$

Student: The theorem says it's invertible.

Instructor: That's right! The μ in this case is 2 and since $\|A\| = 1.6 < 2 = |\mu|$, the hypothesis of the theorem is satisfied and the theorem says that $2I - A$ is invertible. In fact, the determinant of $2I - A$ is $.88 \neq 0$ and we can find the inverse by direct computation.

Instructor: What does this theorem say about the matrix

$$I - A = \begin{pmatrix} .3 & .2 \\ -.5 & -.4 \end{pmatrix}?$$

Student 1: The theorem says it's not invertible.

Instructor: Is that correct?

Student 2: No, the theorem doesn't say anything about this matrix.

Instructor: Student 2 is right! The μ in this case is 1 and since $\|A\| = 1.6 > 1 = |u|$, the hypothesis of the theorem is not satisfied and we cannot tell *from this theorem* if the matrix $I - A$ is invertible or not. In fact, the determinant of $I - A$ is $-.02$ so it *is* invertible. Similarly, for $\mu = 1.2$, the hypothesis of the theorem is not satisfied and we cannot tell *from this theorem* if the matrix $1.2I - A$ is invertible or not. In this case, we find that the determinant of $1.2I - A$ is 0, so the matrix $1.2I - A$ isn't invertible.

Instructor: Now, who can remind us of the definition of eigenvalue and tell some things we know about eigenvalues?

Student: The number λ is an eigenvalue of the matrix A if there is a nonzero vector v for which $Av = \lambda v$. We know that λ is an eigenvalue if and only if $\lambda I - A$ is singular.

Instructor: What does this theorem say about the eigenvalues of A?

Student: If the number λ satisfies $|\lambda| > \|A\|$, then it is not an eigenvalue of A.

Typically, I would then formally state the corollary just given by the student and prove the theorem (by adding up the geometric series) and the corollary. Classes need varying amounts of coaxing and help in such a discussion but they require less coaxing after they are accustomed to the style, especially if the instructor refuses to answer the questions for the class!

A second opportunity for practice is in assigning homework over material not covered in class. Most textbooks include much more material than can be covered in a typical course and much of the material that must be skipped is interesting and valuable. Asking the students to read a section of the book and do some related homework is a reasonable way for them to get practice in reading for understanding. Such material needs to be carefully chosen for this purpose to be sufficiently theoretical to be challenging yet down to earth enough that students can be successful. For example, in some real analysis courses, it might be reasonable to have students read a section in the text defining monotone function and culminating with the proof of the theorem that a monotone function on a finite interval has at most countably many discontinuities. Students could then be asked to prove that the sum of two increasing functions is increasing, give an example of two monotone functions whose sum is not monotone, and perhaps prove that a monotone function on the whole line has at most countably many discontinuities.

Finally, if we really expect students to take learning to read seriously, then we must put it on our tests. Whatever other qualities our students have, they have learned to play the student game well. For their part, the students learn to pass our tests and, for our part, we faculty have capitulated to that hated question "Professor, will this be on the test?" Now we only expect them to learn what we are putting on the test! We can use this to our advantage if we test them on their reading ability.

Clearly, on an hour test, we cannot give the students a book chapter to read and describe. To test reading mathematics that involves the ideas of the course, I have frequently included on an exam a new (to the students) theorem, together with its proof, followed by a question that (I hope) can be answered only by those who read and understood the theorem and proof. The theorems need not be significant, although it's nicer if they are. Moreover, since we want it to be possible to answer the question only by understanding the proof, the conclusion may not be stated in the most elegant or cleanest way. Here are two examples.

The first example is from a course for mathematics education majors. The aims of the course are to develop some understanding of theorems, proofs, and mathematical statements in general. The topics include material on sets, number theory, real numbers, and polynomial equations. This question comes from the first test of the semester which follows the material on number theory; the course does not usually include any discussion of Pythagorean triples, solutions of Diophantine equations, or writing integers as sums of squares.

THEOREM. *If m and n are positive integers such that*

$$2m = n^2 + 1,$$

then m is the sum of the squares of two integers.

Proof. Since $n^2 + 1$ is even, n^2 and therefore n must be odd. That is, there is an integer k such that $n = 2k + 1$. This means that

$$2m = n^2 + 1 = (2k + 1)^2 + 1 = (4k^2 + 4k + 1) + 1 = 4k^2 + 4k + 2$$

It follows that

$$m = 2k^2 + 2k + 1 = k^2 + (k^2 + 2k + 1) = k^2 + (k + 1)^2$$

which expresses m as the sum of the squares of two integers. ■

Problem:

$$2 \cdot 3785 = 7570 = 87^2 + 1$$

Write 3785 as the sum of the squares of two integers.

Of course, if we were really going to state and prove this "theorem" in a book, we would probably spell out how m is written as the sum of two squares, but to do so in this context would obviate the need to understand the proof. What I expect from this question is that students will see from the proof that for $k = (n - 1)/2$, we have $m = k^2 + (k + 1)^2$. To write 3785 as the sum of two squares, I expect them to see that in the notation of the theorem, $n = 87$, so $k = 43$ and by the proof of the theorem $3785 = 43^2 + 44^2$. If students perform this calculation, I infer that they read and understood the proof well enough to do so.

The second example comes from a linear algebra course for engineers. At the time of the test, the students had studied the spectral mapping theorem for polynomials which implies that the positive integer powers of a diagonalizable matrix are similar to the powers of the diagonal matrix, but had not seen the theorem on square roots of positive definite matrices.

THEOREM. *If A is a diagonalizable matrix all of whose eigenvalues are nonnegative, then there is a matrix B with nonnegative eigenvalues such that $B^2 = A$.*

Proof. Since A is diagonalizable, there is an invertible matrix S such that $L = S^{-1}AS$ is diagonal. The diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ of L (which are the eigenvalues of A) are nonnegative by hypothesis. Let $\mu_j = \sqrt{\lambda_j}$ be the nonnegative square roots of the eigenvalues and let M be the diagonal matrix with diagonal entries $\mu_1, \mu_2, \ldots, \mu_n$. Clearly, $M^2 = L$. Since similar matrices have the same eigenvalues, the matrix $B = SMS^{-1}$ has eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$, which are nonnegative. Moreover,

$$B^2 = (SMS^{-1})^2 = SMS^{-1}SMS^{-1} = SM^2S^{-1} = SLS^{-1} = A. \quad ■$$

Problem: The matrix $A = \begin{pmatrix} 10 & -9 \\ 6 & -5 \end{pmatrix}$ has eigenvalues 1 and 4. Find a matrix S as above and use it to find a matrix B with positive eigenvalues such that $B^2 = A$.

In this question, I expect the students to understand from the proof of the theorem that the problem is to be solved by diagonalizing A and solving the corresponding problem for the diagonal matrix. In particular, I expect them to use the given eigenvalues of A to find a basis of eigenvectors, and thereby to construct S, L, M, and B.

I have not done a comparative analysis of my classes to see if trying to teach reading makes a difference in how much students learn, but I am convinced it helps. At the very least, if we communicate to the students that we think it is reasonable that they read mathematics and understand it and that we expect it of them, then they may come to expect it of themselves. We will all be better off if the students view the textbook as a source of information, not just as a list of exercises interspersed with messages for the instructor.

I'd like to thank Dick Hunt, J. J. Price, and Bob Zink for their suggestions and support and the referee for suggestions concerning this note.

bisector of the angle BAC, F is the foot of the perpendicular from A to BC, and AG is perpendicular to AE (i.e., AG bisects one of the exterior angles at A).

Prove that $AB \cdot AC = DF \cdot EG$.

E 3418. *Proposed by E. T. Parker, University of Illinois at Urbana-Champaign.*

For $k = 1, 2, \ldots, 9$ let S_k be the set of positive integers n such that the number of digits in the decimal expansion of n is a multiple of 10 and such that each of the digits $1, \ldots, k$ occurs in exactly one-tenth of the places. For which values of k does

$$\sum_{n \in S_k} n^{-1}$$

converge?

E 3419. *Proposed by Marcin E. Kuczma, University of Warsaw, Poland.*

Let $F(n)$ be the number of n by n matrices with entries 0 or 1 and row and column sums equal to 2. Let $f(n) = n^{1/2}F(n)(n!)^{-2}$. Prove that $\lim_{n \to \infty} f(n)$ exists and has a value between 0 and 1.

E 3420. *Proposed by S. G. Merzlyakov, Mathematical Institute, Ural Branch of the Academy of the USSR.*

Find all functions f holomorphic on the square $S = \{x + iy: -1 < x < 1, -1 < y < 1\}$ in the complex plane for which there exist real-valued functions α and β on $(-1, 1)$ satisfying

$$|f(x + iy)| = \alpha(x) + \beta(y)$$

throughout S.

SOLUTIONS OF ELEMENTARY PROBLEMS

Inverting a Transformation by Equilateral Triangles

E 3257 [1988, 259]. *Proposed by I. A. Sakmar, Istanbul, Turkey.*

Let P, Q, R be the new vertices of equilateral triangles constructed outwardly on the edges of a given triangle ABC. (Cf. H. S. M. Coxeter, Introduction to Geometry, New York, 1961, p. 22.)

(a) Show that any triangle PQR which can be obtained in this way arises from a unique triangle ABC, and give a construction for recovering triangle ABC from triangle PQR.

(b) Show that not every triangle PQR can be so obtained.

Solution by J. G. Mauldon, Amherst College, Amherst, Massachusetts. (a) Let U, V, W be the new vertices of equilateral triangles constructed *inwardly* on the edges of the triangle PQR. Then A, B, C are necessarily the midpoints D, E, F of the sides of the triangle UVW.

(b) The triangle $\triangle PQR$ with sides of length p, q, r can be so obtained if and only if $10 \sin P \sin Q \sin R > \sqrt{3}(\sin^2 P + \sin^2 Q + \sin^2 R)$ or, equivalently, if and only if $11(p^2 + q^2 + r^2)^2 > 25(p^4 + q^4 + r^4)$. In particular, if $\{p, q, r\} = \{2, 4, 5\}$ or $\{4, 5, 8\}$ the triangle $\triangle PQR$ is unobtainable and an isosceles $\triangle PQR$ with base angle α is obtainable if and only if $1 < \sqrt{3} \tan \alpha < 9$.

Proof of (*a*). We take the plane of discourse to be the complex plane $\mathbb{C}$ and we write $\omega = e^{2i\pi/3}$, so that $1 + \omega + \omega^2 = 0 = \omega^3 - 1$. By a *triangle* we shall mean an ordered nondegenerate triangle. We identify each point of $\mathbb{C}$ with its (single, complex) coordinate, so that 0 is the origin O and the triangle $\triangle XYZ$ is a positively oriented equilateral triangle if and only if (for some $\lambda, \mu \in \mathbb{C}$) $\langle X, Y, Z \rangle = \langle \lambda + \mu, \lambda + \mu\omega, \lambda + \mu\omega^2 \rangle$ or, equivalently, if and only if $X + \omega Y + \omega^2 Z = 0$.

We assume without loss of generality that $\triangle ABC$ is positively oriented, so that the requirement that equilateral triangles are constructed outwardly on the sides of $\triangle ABC$ implies that $\triangle PCB, \triangle QAC, \triangle RBA$ are positively oriented equilateral triangles. We shall deduce below in (*) that then $\triangle PQR$ is positively oriented, so that similarly the requirement that equilateral triangles are constructed inwardly on the sides of $\triangle PQR$ implies that $\triangle UQR, \triangle VRP, \triangle WPQ$ are positively oriented equilateral triangles. Hence all six triangles $\triangle PCB, \triangle QAC, \triangle RBA, \triangle UQR, \triangle VRP, \triangle WPQ$ are positively oriented equilateral triangles, and consequently we have

$$\begin{aligned} P &= -\omega C - \omega^2 B, \quad U = -\omega Q - \omega^2 R, \quad D = (V + W)/2, \\ Q &= -\omega A - \omega^2 C, \quad V = -\omega R - \omega^2 P, \quad E = (W + U)/2, \\ R &= -\omega B - \omega^2 A; \quad W = -\omega P - \omega^2 Q; \quad F = (U + V)/2. \end{aligned}$$

By substitution we obtain, as required, $\langle D, E, F \rangle = \langle A, B, C \rangle$, completing the proof of (a).

Proof of (*b*). For any triangle $\triangle XYZ$ we denote by $\mathbf{S}(\triangle XYZ)$ the sum of the squares of the lengths of its sides and we write $\mathbf{T}(\triangle XYZ) = (4\sqrt{3})$(signed area of $\triangle XYZ$), where this area is regarded as positive or negative according as $\triangle XYZ$ is positively or negatively oriented. We also write $\phi_k(\triangle XYZ) = |X + \omega^k Y + \omega^{2k} Z|^2$ $(k = 1, 2)$, observing that ϕ_k is unchanged under any translation or rotation (in $\mathbb{C}$) of $\triangle XYZ$ so that, if $\triangle XYZ \cong \triangle X'Y'Z'$ (directly congruent), then, for $k \in \{1, 2\}$, $\phi_k(\triangle XYZ) = \phi_k(\triangle X'Y'Z')$.

Two interesting general identities are $\phi_2(\triangle XYZ) - \phi_1(\triangle XYZ) = \mathbf{T}(\triangle XYZ)$ and $\phi_1(\triangle XYZ) + \phi_2(\triangle XYZ) = \mathbf{S}(\triangle XYZ)$. In proving these, the above paragraph shows that we lose no generality in taking $\langle X, Y, Z \rangle = \langle re^{i\theta}, 0, s \rangle$, with $r, s > 0$. Then

$$\begin{aligned} \phi_2(\triangle XYZ) - \phi_1(\triangle XYZ) &= (re^{i\theta} + \omega s)(re^{-i\theta} + \omega^2 s) - (re^{i\theta} + \omega^2 s)(re^{-i\theta} + \omega s) \\ &= (e^{4i\pi/3} - e^{-4i\pi/3})rs(e^{i\theta} - e^{-i\theta}) \\ &= 2i \sin(4\pi/3)rs(2i \sin\theta) \\ &= (2\sqrt{3})rs \sin\theta = \mathbf{T}(\triangle XYZ). \end{aligned}$$

The other result can be obtained similarly or, for general $\triangle XYZ$, by direct computation.

Turning to the specifics of the present problem, we find $P + \omega Q + \omega^2 R = A + \omega B + \omega^2 C$ and $P + \omega^2 Q + \omega R = -2(A + \omega^2 B + \omega C)$, and consequently

$\phi_1(\triangle PQR) = \phi_1(\triangle ABC)$ and $\phi_2(\triangle PQR) = 4\phi_2(\triangle ABC)$. Substituting these results in the above two general identities, we find

$$2\mathbf{T}(\triangle PQR) = 5\mathbf{T}(\triangle ABC) + 3\mathbf{S}(\triangle ABC),$$
$$8\mathbf{T}(\triangle ABC) = 5\mathbf{T}(\triangle PQR) - 3\mathbf{S}(\triangle PQR).$$

The first equation shows, as required in (a) above, that

$$\mathbf{T}(\triangle ABC) > 0 \quad \Rightarrow \quad \mathbf{T}(\triangle PQR) > 0. \qquad (*)$$

The second equation shows that $\triangle PQR$ is obtainable iff $5\mathbf{T}(\triangle PQR) > 3\mathbf{S}(\triangle PQR)$.

Then Heron's formula (*op. cit.* p. 12, 1.53) for the area of a triangle yields the second of the two given conditions, and the first is obtained from

$$5(2\sqrt{3})PQ \cdot PR \sin(\angle QPR) > 3(QR^2 + RP^2 + PQ^2)$$

together with $QR : RP : PQ :: \sin P : \sin Q : \sin R$.

Editorial comment. A closely related question is asked and answered elegantly (using motions) in a paper of H. G. Steiner, "Bewegungsgeometrische Lösung einer Dreieckskonstruktion," *Math.-Phys. Semesterber.* 5 (1956), 132–137, as is noted by John E. Wetzel in "Some converses of Napoleon's Theorem", a paper forthcoming in this MONTHLY. The problem is studied in greater detail in another paper of Wetzel, "An elaboration on an example of H. G. Steiner," *Math. Semesterber.* 37 (1990), 88–95. See also J. G. Mauldon, "Similar triangles," this MONTHLY 39 (1966), 165–174. Most solvers proceeded by constructing *outward* equilateral triangles on the sides of PQR, obtaining points X, Y, Z, and proving that A, B, C, are the midpoints of PX, QY, RZ, respectively, a construction that was given in 1869 by L. Kiepert in "Question 864," *Nouv. Ann. Math.* (2)8(1869), 40–42, as Wetzel points out in his papers.

The mapping $A, B, C \to P, Q, R$ is the restriction to positively oriented triples ABC of a non-singular linear mapping of $\mathbb{C}^3$ onto $\mathbb{C}^3$. (The determinant of this mapping is -2.) Accordingly both the uniqueness assertion in (a) and the existence of the inverse mapping in (b) (to triples of unspecified orientation) are immediate. The image of the positively oriented triples ABC under this mapping can be described in various ways, including the two given above; a geometric description may be found on page 94 of Wetzel's paper in *Math. Semesterber.*

Solved also by R. L. Bishop, J. Dou (Spain), M. Golomb, H. Guggenheimer, J. Heuver (Canada), E. Kitchen, V. Konecny, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), J. H. Meyer, R. D. Nelson (England), C. Petalas & T. Vidalis (Greece), R. A. Simon (Chile), S. Stahl, and the proposer.

The Maximum Value of a Symmetric Function on Points of a Simplex

E 3263 [1988, 351]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

For $n \geqslant 2$, let $H_n = \{X = (x_1, \ldots, x_n)\colon \Sigma x_i = 1, x_i > 0 \text{ for all } i\}$. For $1 \leqslant k \leqslant n$ and $X \in H_n$, let

$$S_k(X) = \sum \frac{x_{i_1}}{1 - x_{i_1}} \cdots \frac{x_{i_k}}{1 - x_{i_k}},$$

where the sum extends over all k-tuples $(i_1, \ldots, i_k)$ with $1 \leqslant i_1 < \cdots < i_k \leqslant n$.

Put

$$M_k(n) = \sup_{X \in H_n} S_k(X).$$

It is not hard to see that $M_n(n) = (n-1)^{-n}$.

(a) Show that $M_2(n) = 1$.

(b) Show that $M_3(4) = 4/27$.

(c)* For what pairs k, n with $3 \leqslant k \leqslant n$ is it true that $M_k(n) = \binom{n}{k}(n-1)^{-k}$?

Composite solution by J. A. M. Potters, Catholic University Nijmegen, Toernooiveld, The Netherlands, Woo Jung Min (student), Yonsei University, Seoul, Korea, and the editors. The answer to (c) is "for all such pairs". We examine these maximizations in the full simplex $\bar{H}_n$ with boundary, where the variables may vanish. We show that the value of S_k is maximized when all n variables are equal if $k > 2$, and where all but 2 variables vanish if $k = 2$.

We will argue that we need look for the optimum only among vectors having all non-zero values equal. If so, then $M_k(n) = \max_{k \leqslant j \leqslant n}\binom{j}{k}(j-1)^{-k}$, where j represents the number of non-zero variables at the possible local extrema. For $k = 2$, this function is $j/(2j-2)$, which decreases in j; the maximum occurs at $j = 2$. For all other k, the function increases with j, so the maximum occurs at $j = n$ and equals $\binom{n}{k}(n-1)^{-k}$; the increasing property follows readily from the identity

$$\binom{j}{k}(j-1)^{-k} = \frac{1}{k!}\left(1 + \frac{1}{j-1}\right)\prod_{i=1}^{k-2}\left(1 - \frac{i}{j-1}\right).$$

Suppose that more than two of the variables are non-zero. If x_1 and x_2 are two such variables, we can write

$$S_k(X) = \frac{x_1}{1-x_1}\frac{x_2}{1-x_2}U(\bar{X}) + \left(\frac{x_1}{1-x_1} + \frac{x_2}{1-x_2}\right)V(\bar{X}) + W(\bar{X}),$$

where $\bar{X} = (x_3, \ldots, x_n)$. We can use Lagrange multipliers to maximize this function of x_1, x_2 over the region where $x_1 + x_2 = s$, $0 < s < 1$. Differentiating $S_k(x_1, x_2) - \lambda(x_1 + x_2 - s)$ with respect to x_1 and x_2 yields the following two equations:

$$\frac{1}{(1-x_1)^2}\left(\frac{x_2}{1-x_2}U + V\right) = \lambda$$

$$\frac{1}{(1-x_2)^2}\left(\frac{x_1}{1-x_1}U + V\right) = \lambda.$$

Subtracting these yields

$$[x_2(1-x_2) - x_1(1-x_1)]U = \left[(1-x_1)^2 - (1-x_2)^2\right]V,$$

which is equivalent to

$$(x_1 - x_2)(x_1 + x_2 - 1)U = (x_1 - x_2)(x_1 + x_2 - 2)V.$$

If $x_1 \neq x_2$, then this reduces to $(1-s)U = (2-s)V$. Letting $p = x_1x_2$, we can

eliminate U in the expression for S at the maximum to obtain

$$S = \frac{p(2-s)}{(1-s+p)(1-s)}V + \frac{s-2p}{1-s+p}V + W = \frac{s}{1-s}V + W.$$

This is independent of p and the same value is obtained at the point $(x_1 + x_2, 0, \overline{X})$. This implies that for any maximizing (or minimizing) point with more than two non-zero coordinates not all of which are equal, there is another point with fewer non-zero coordinates on which the value of S is the same. Hence S must be maximized (or minimized) at a point where all non-zero coordinates are equal.

Note that if only one coordinate of X is non-zero, then $S_k(X)$ should be regarded as 0 if $k > 2$ and 1 if $k = 2$. If the variables are not permitted to vanish, then for $n > 2$ and $k = 2$ the maximum value of 1 can only be approached and not attained.

Because the value of S on points with equal non-zero coordinates is monotonic in the number of non-zero coordinates, the argument given yields all extreme points, minima as well as maxima. The minima occur when there are fewer than k non-vanishing variables if $k > 2$, and when there are n equal variables if $k \leqslant 2$.

No other solutions were received, except for a solution of (b) by David Secrest.

Feuerbach When Ready

E 3293 [1988, 873]. *Proposed by Joseph Keane (student), Carnegie-Mellon University, Pittsburgh, and Gregg Patruno, The First Boston Corporation, New York, NY.*

Suppose that the distinct circles C_1 and C_2 intersect at P and Q. Suppose that the tangent to C_1 at P intersects C_2 again at A, the tangent to C_2 at P intersects C_1 again at B, and the line AB separates P and Q. Let C_3 be the circle externally tangent to C_1, externally tangent to C_2, tangent to line AB, and lying on the same side of AB as Q. Prove that the circles C_1 and C_2 intercept equal segments on one of the tangents to C_3 through P.

Solution by Richard K. Guy, University of Calgary, Calgary, Alberta, Canada. Invert with respect to a circle centered at P. The circles C_1, C_2 invert into straight lines $Q'B', Q'A'$ (Q' is the inverse of Q, etc.) and the tangents at P into lines through P, parallel to $Q'B', Q'A'$. The line AB inverts into the circle $PA'B'$, and "AB separating P and Q" is equivalent to "Q' lying inside circle $PA'B'$." So we have a parallelogram $PA'Q'B'$, obtuse-angled at P and Q'. Draw the line through P, parallel to $A'B'$, meeting $Q'B', Q'A'$ in R, S. Then P, A', B' are the midpoints of the sides of triangle $Q'RS$, and RS is the inverse of the line through P on which C_1, C_2 intercept equal segments. The circle $PA'B'$ is the nine-point circle of triangle $Q'RS$; by Feuerbach's Theorem, it touches the incircle of that triangle. But this incircle is the inverse of the circle C_3 in the original figure, touching C_1, C_2 (externally) and the line AB.

Editorial comment. The proposers and several solvers remarked that the statement of the problem can be slightly generalized, using the fact that the nine-point circle of a triangle is tangent not only to the incircle but also to the three excircles.

Feuerbach's Theorem is proved in §5.6 of H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, New Mathematical Library #19, Mathematical Association of America or on pp. 9–10 of D. Pedoe, *Circles*, Pergamon, New York, 1957.

Solved also by J. Anglesio (France), J. Dalbec, J. Fukuta (Japan), W. Janous (Austria), O. P. Lossers (The Netherlands), A. Nijenhuis, H. Steinacker (Austria), Lamar University Problem Group, and the proposers.

A Ratio on the Newton Line

E 3299 [1988, 954]. *Proposed by Jihad Yamout, Ohio University, Athens.*

Suppose $ABCD$ is a plane quadrilateral with no two sides parallel. Put $E = \overleftrightarrow{AB} \cap \overleftrightarrow{CD}$ and $F = \overleftrightarrow{AD} \cap \overleftrightarrow{BC}$. If M, N, P are the midpoints of AC, BD, EF, respectively, and $\overrightarrow{AE} = a\overrightarrow{AB}$, $\overrightarrow{AF} = b\overrightarrow{AD}$, where a and b are nonzero real numbers, prove that $\overrightarrow{MP} = ab\overrightarrow{MN}$.

Solution by Hans Kappus, Rodersdorf, Switzerland. With respect to a suitably chosen affine coordinate system, we may assume that $A = (0,0)$, $B = (1,0)$, $C = (x,y)$, and $D = (0,1)$. Then $M = (x/2, y/2)$, $N = (1/2, 1/2)$, $E = (a,0)$, $F = (0,b)$, and $P = (a/2, b/2)$. Because of the colinearity of D, C, E and of B, C, F respectively, x, y can now be expressed in terms of a, b as $x = a(1-b)/(1-ab)$ and $y = b(1-a)/(1-ab)$. Consequently,

$$\overrightarrow{MN} = \frac{1}{2}(1-x, 1-y) = \frac{1}{2(1-ab)}(1-a, 1-b)$$

$$\overrightarrow{MP} = \frac{1}{2}(a-x, b-y) = \frac{ab}{2(1-ab)}(1-a, 1-b) = ab\overrightarrow{MN}.$$

Editorial comment. H. Guggenheimer noted that the colinearity of M, N, P was proved by Gauss (*Works*, vol. 4, p. 385–392) using Corollary III to Lemma XXV of Newton's *Principia*. The line determined by M, N, P is sometimes called the Newton line.

Solved also by J. C. Binz (Switzerland), S. Gaignoux (France), H. Guggenheimer, T. Hermann (Hungary), J. Heuver (Canada), P. L. Hon (Canada), W. Janous (Austria), R. H. Jeurissen (The Netherlands), P. Khajeh-Khalili, L. Kuipers (Switzerland), H. Lipman, O. P. Lossers (The Netherlands), A. Nijenhuis, A. Pedersen (Denmark), I. A. Sakmar (Turkey), A. Sarkar (England), V. Schindler (East Germany), M. Vowe (Switzerland), J. B. Wilker (Canada), Lamar University Problem Solving Group, the Stanford Ordinal, and the proposer.

A Characterization of Consecutive Composites

E 3308 [1989, 154]. *Proposed by Ion Cucurezeanu, Liceul 10 Constanta, Romania.*

Suppose n is a positive integer greater than 4. Prove that the following two conditions are equivalent:

(i) both n and $n+1$ are composite,

(ii) the integer nearest to $(n-1)!/(n^2+n)$ is even.

Solution and extension by Jorge-Nuno O. Silva and Owen Brison, Faculdade de Ciencias de Lisboa, Lisbon, Portugal. We prove the following more general result. If p is an arbitrary prime and $n \geqslant 2p + 1$ (or $n \geqslant 10$ if $p = 3$), the following are equivalent:

(i) both n and $n + 1$ are composite,

(ii) the integer nearest to $(n - 1)!/(n^2 + n)$ is a multiple of p.

If n is a positive integer and q is prime, let $|n|_q$ denote the largest integer μ such that $q^\mu | n$. We claim the following:

LEMMA. *$|m!|_q \geqslant |r|_q$ for integers m, q, r such that q is prime and $\max\{4, q\} \leqslant m < r \leqslant m + q$. Furthermore, the inequality is strict for $m \geqslant 2q$ (unless $q = 3$, in which case it is strict for $m \geqslant 9$).*

For the proof of the lemma, let k be the largest integer such that $q^k \leqslant m$; this implies $|m!|_q \geqslant k$, and $m \geqslant q$ implies $k \geqslant 1$. Hence we may assume that r is the unique multiple of q in $(m, m + q]$, which implies $1 \leqslant |r|_q \leqslant k + 1$. If $m \geqslant 3q$, then at least three distinct multiples of q appear in the product for $m!$, including q^k; this implies $|m!|_q \geqslant k + 2 > |r|_q$. If $q \leqslant m < 2q$, then $r = 2q$; with $m \geqslant 4$, this implies $q > 2$, which implies $|r|_q = 1 \leqslant k \leqslant |m!|_q$. Finally, if $2q \leqslant m < 3q$, then $r = 3q$. Since $q^2 | m!$, we have $|m!|_q \geqslant 2$, and $|r|_q = 1$ if $q \neq 3$. If $q = 3$, then $|r|_q = 2$, but then $m < 9$ and the claim still holds. This completes the proof of the lemma.

We prove the main equivalence by considering three cases for the pair $\{n, n + 1\}$, always assuming $n \geqslant 2p + 1$ (or $n \geqslant 10$ if $p = 3$).

Case I, n and $n + 1$ are both composite. It suffices to show that $R_n = (n - 1)!/(n^2 + n)$ is an integer divisible by p. We apply the Lemma with $m = n - 1$, q being successively p or the prime factors of n and $n + 1$, and r being n or $n + 1$. We find that $(n - 1)!$ has more factors of p than $n(n + 1)$ (since $\gcd(n, n + 1) = 1$) and as many factors of any other prime as $n(n + 1)$. Thus R_n has the desired property, i.e., is an integer divisible by p.

Case II, n is prime. It suffices to show that $S_n = \{(n - 1)! + n + 1\}/(n^2 + n)$ is an integer not divisible by p. By Wilson's Theorem, $[(n - 1)! + 1]/n$ is an integer and so $\{(n - 1)! + n + 1\}/n$ is an integer. Note that $n + 1 = 2 \cdot [(n + 1)/2]$ divides $(n - 1)!$ when n is prime and $n \geqslant 5$. By the Lemma with $m = n - 1$ and $r = n + 1$, we have $|m!|_p > |r|_p$; this implies that $(n - 1)!/(n + 1)$ is an integer divisible by p. Now we rewrite S_n as $\{(n - 1)!/(n + 1) + 1\}/n$. The numerator is an integer not divisible by p. However, since $\gcd(n, n + 1) = 1$, the numerator is still a multiple of n. Thus S_n has the desired property, i.e., it is an integer not divisible by p.

Case III, $n + 1$ is prime. It suffices to show that $T_n = \{(n - 1)! + n\}/(n^2 + n)$ is an integer not divisible by p. By Wilson's Theorem, $(n! + 1)/(n + 1)$ is an integer. Since $(n - 1)! + 1 - (n! + 1)/(n + 1) = \{(n - 1)! + n\}/\{n + 1\}$, it follows that $[(n - 1)! + n]/(n + 1)$ is also an integer. By the Lemma with $m = n - 1$ and $r = n$, we have $|m!|_p > |r|_p$; this implies that $(n - 1)!/n$ is an integer divisible by p. Now we rewrite T_n as $\{(n - 1)!/n + 1\}/(n + 1)$. The numerator is an integer not divisible by p. However, since $\gcd(n, n + 1) = 1$, the numerator is still a multiple of $n + 1$. Thus T_n has the desired property, i.e., it is an integer not divisible by p.

Editorial comment. The above proof simplifies somewhat when $p = 2$. In fact one can use the identities

$$R_n = (n-1)!/n + n!/(n+1) - (n-1)! \tag{I}$$

$$S_n = \{(n-1)!+1\}/n + n!/(n+1) - (n-1)! \tag{II}$$

$$T_n = (n!+1)/(n+1) + (n-1)!/n - (n-1)! \tag{III}$$

in the three cases considered above, where each term on the right is an integer whose parity can be readily determined.

Solved also by the proposer and 34 others.

An Arcsine Inequality

E 3326 [1989, 445]. *Proposed by George Baloglou, University of Kansas, Lawrence.*

If $0 < \theta < \pi$, prove that

$$\arcsin\frac{\theta}{6} + \arcsin\left(\frac{2}{3}\sin\frac{\theta}{4}\right) < \frac{\theta}{3}.$$

Solution by Jesús Ferrer, Oliva, Valencia, Spain. Set $x = \theta/4$, let $I = \{x: 0 < x < \pi/4\}$, and consider the function $f(x) = \arcsin(2x/3) + \arcsin(2(\sin x)/3) - 4x/3$. We will show $f(x) < 0$ on I. Since $f(0) = 0$, it suffices to show $f'(x) < 0$ on I. From $f'(x) = 2[(9 - 4x^2)^{-1/2} + (9 + 5\tan^2 x)^{-1/2} - 2/3]$, it suffices to show $g(x) < 2/3$ on I, where $g(x) = (9 - 4x^2)^{-1/2} + (9 + 5\tan^2 x)^{-1/2}$. Since $g(0) = 2/3$, it suffices to show $g'(x) < 0$ on I. Since $g'(x) = 4x(9 - 4x^2)^{-3/2} - 5\tan x(1 + \tan^2 x)(9 + 5\tan^2 x)^{-3/2}$, it suffices to prove

$$A = \frac{(9 + 5\tan^2 x)^{3/2}}{\tan x(1 + \tan^2 x)} < \frac{5(9 - 4x^2)^{3/2}}{4x} = B.$$

Now $\sin(x)/x > \sin(\pi/4)/(\pi/4) = 2\sqrt{2}/\pi$ implies $1 + \tan^2 x = (1 - \sin^2 x)^{-1} > \pi^2(\pi^2 - 8x^2)^{-1}$ on I. Hence

$$A = \{5 + 4/(1 + \tan^2 x)\}(5 + 9/\tan^2 x)^{1/2} < (9\pi^2 - 32x^2)^{3/2}/(2\sqrt{2}\pi^2 x).$$

We need only prove that the last expression is at most B, which is equivalent to $(9\pi^2 - 32x^2)/(9 - 4x^2) < (25\pi/2)^{1/3}\pi$. This follows from the fact that the left side of this final inequality is an increasing function whose value at $x = \pi/4$ is $28\pi^2/(36 - \pi^2)$, which is less than $(25\pi/2)^{1/3}\pi$.

Solved also by J. Anglesio (France), J.-M. Becker (France), J. S. Frame, M. E. Kuczma (Poland), K.-W. Lau (Hong Kong), H. Lipman, K. McInturff, J.-M. Monier (France), and the proposer (twice alone and once with S. Reyner). Six incorrect solutions were received.

Missing Only the kth Powers

E 3340 [1989, 642]. *Proposed by Courtney Moen, U.S. Naval Academy, Annapolis, MD.*

Suppose k is a given integer $\geqslant 2$. For $n \in \mathbb{N}$ (the set of positive integers) put

$$f(n) = \left\lfloor (n + n^{1/k})^{1/k} \right\rfloor + n,$$

where $\lfloor\ \rfloor$ denotes the greatest integer function. What is the range of the function f?

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. The range of f consists of all positive integers that are not perfect kth powers.

First we show that non-kth powers are in the range of f. Since $k \geq 2$, the binomial theorem implies that $s^k + s + 1 \leq (s+1)^k$ for all $s \geq 0$. With $s = a - 1$, we have $(a-1)^k \leq a^k - a$. For $a, m \in \mathbb{N}$ with $a^k < m < (a+1)^k$, we claim that $m = f(m-a)$. Let $n = m - a$, so that $a^k - a < n < m$. Our various inequalities now imply

$$a^k \leq (m-a) + (a-1) < n + n^{1/k} < (m-a) + (a+1) \leq (a+1)^k.$$

Thus $f(n) = a + (m-a) = m$.

Next we show that the kth powers are missed. Note that f is a strictly increasing function, so that $f(1) = 2$ implies 1 is missed. For any larger a^k, it suffices to present two consecutive values whose images are $a^k - 1$ and $a^k + 1$. If m is not a kth power, we showed above that m is the image of $m - \lfloor m^{1/k} \rfloor$. Hence $f(a^k + 1 - a) = a^k + 1$ and $f(a^k - a) = f(a^k - 1 - (a-1)) = a^k - 1$.

Solved also by J. Balogh (student, Hungary), J. C. Binz (Switzerland), J. Duemmel, N. J. Fine, R. J. Hendel, W. Janous (Austria), H. Ki (South Korea), K. -W. Lau (Hong Kong), M. Leeney (Ireland), O. P. Lossers (The Netherlands), J. H. Nieto (Venezuela), A. Nijenhuis, H. Noland, V. V. Pambuccian, A. Pedersen (Denmark), B. M. de Weger (The Netherlands), R. N. Will, J. Zhang (China), Arizona State Problems Group, National Security Agency Problems Group, and the proposer. There was one incorrect solution.

ADVANCED PROBLEMS

6646. *Proposed by Daniel Goffinet, St. Etienne, France.*

Suppose f is a continuous function from $[0,1] \times [0,1]$ to R such that

$$\left(\int_0^1 \left\{ \int_0^1 f(x,y)\,dx \right\}^2 dy \right)^{1/2} = \int_0^1 \left(\int_0^1 \{f(x,y)\}^2\,dy \right)^{1/2} dx.$$

Prove that there exist continuous functions u and r from $[0,1]$ to R such that $\int_0^1 \{u(y)\}^2\,dy = 1$, $r(x) \geq 0$ for all x in $[0,1]$, and $f(x,y) = r(x)u(y)$ for all x and y in $[0,1]$.

6647. *Proposed by A. J. Vince, Bristol Polytechnic, Bristol, England.*

Let $S_n = \{1, \zeta, \zeta^2, \ldots, \zeta^{n-1}\}$ be the set of n-th roots of unity and suppose f is any function on S_n into the set of complex numbers of absolute value one. For every positive integer k less than $n/2$ prove that there exist integers i and j such that

$$|\zeta^i - \zeta^j| \geq |1 - \zeta^k| \geq |f(\zeta^i) - f(\zeta^j)|.$$

6648. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Let Ω be the region obtained by removing the points $0, 1, \infty$ from the Riemann sphere. Find all nonconstant holomorphic functions defined on Ω which map Ω into itself.

SOLUTIONS OF ADVANCED PROBLEMS

Eigenvalues of Sine and Cosine Matrices

6591 [1989, 66]. *Proposed by Ronald J. Evans, University of California, San Diego, and Jerrold R. Griggs, University of South Carolina, Columbia.*

Let n be an integer exceeding 4 and put $t = \lfloor (n-1)/2 \rfloor$. Find the eigenvalues of the following t by t matrices:

(i) the matrix S in which the element in the jth row and kth column is $\sin(2\pi jk/n)$, $1 \le j, k \le t$.

(ii) the matrix C in which the element in the jth row and kth column is $\cos(2\pi jk/n)$, $1 \le j, k \le t$.

Solution by David Callan, University of Bridgeport, Bridgeport, CT, and Grace Wahba, University of Wisconsin, Madison (jointly). Let λ denote $\sqrt{n}/2$. The eigenvalues of S are λ, with multiplicity $\lceil t/2 \rceil$, and $-\lambda$, with multiplicity $\lfloor t/2 \rfloor$. The matrix C has simple eigenvalues -1 and 0 if $n \equiv 0 \bmod 4$, $\pm 1/\sqrt{2}$ if $n \equiv 2 \bmod 4$, and $-1/2$ if n is odd. If the remaining eigenvalues of C are even in number they comprise an equal number of λ and $-\lambda$ values; otherwise, there is one more λ.

To show this, first calculate S^2 and C^2. Let I denote the t by t identity matrix, let J denote the t by t matrix with all entries $1/2$, and let K denote the t by t matrix in which the entry in the jth row and kth column is $\{1 + (-1)^{j+k}\}/2$. Then by elementary trigonometric identities

$$S^2 = (n/4)I$$

while

$$C^2 = (n/4)I - J$$

for n odd and

$$C^2 = (n/4)I - K$$

for n even. The eigenvalues of C^2 are now easily found, and are given in the second column of the table below (multiplicities in parentheses). This determines the desired eigenvalues up to sign, and a determination of the traces of C and S is sufficient to settle the matter. These are obtained from the value of

$$\sum_{j=1}^{t} \exp\left(\frac{2\pi i j^2}{n}\right) = \operatorname{Tr}(C) + i \operatorname{Tr}(S)$$

which is known from the theory of Gaussian sums. (Cf. E. Landau, *Vorlesungen über Zahlentheorie*, Satz 211, or T. Nagell, *Introduction to Number Theory*, Theorem 99.) Thus the result follows from the values provided in the third and fourth columns of the table below.

n mod 4	Eigenvalues of C^2	Tr(C)	Tr(S)
0	$0, 1, \frac{n}{4}((t-2)\text{times})$	$\lambda - 1$	λ
1	$\frac{1}{4}, \frac{n}{4}((t-1)\text{times})$	$\lambda - \frac{1}{2}$	0
2	$\frac{1}{2}(\text{twice}), \frac{n}{4}((t-2)\text{times})$	0	0
3	$\frac{1}{4}, \frac{n}{4}((t-1)\text{times})$	$-\frac{1}{2}$	λ

Callan and Wahba also mention that an analogous problem involving a matrix of exponentials (known as the finite Fourier transform matrix), namely $E = (e_{jk})$ where

$$e_{jk} = \exp\left(\frac{2\pi ijk}{n}\right), \qquad (1 \le j, k \le n),$$

appeared here as Problem 3824 [1937, 251; 1939, 304]. Griggs remarks that he came across matrices of this type while examining certain random walk problems.

Solved also by the proposers.

A Property of Two Dimensions

6593 [1989, 165]. *Proposed by Michel Balazard, Limoges, France.*

Suppose E is a two-dimensional vector space of continuous real-valued functions on $[0, 1]$ such that for every $x \in [0, 1]$ there exists g in E with $g(x) \neq 0$.

Suppose S is a nonempty subset of E such that

$$G(x) = \sup_{g \in S} g(x) < +\infty$$

for every $x \in [0, 1]$. Show that G is continuous and give an example to show that the result fails if E is three-dimensional instead of two-dimensional.

Solution by Christopher P. Grant, University of Utah, Salt Lake City. For a counterexample in the three-dimensional case take E to be the span of $\{1, 1 + x, 1 + \sqrt{x}\}$ and S to be the set

$$\{-\lambda^2 x + \lambda\sqrt{x} : \tfrac{1}{2} \le \lambda < \infty\}.$$

A routine calculation gives $G = 0$ at $x = 0$ and $G = \frac{1}{4}$ on the rest of $[0, 1]$. Note that this counterexample also shows that (in the two-dimensional case) the non-vanishing condition is appropriate.

We now prove that, with the stated hypotheses, G is continuous at every x_0 in $[0, 1]$ if E is two-dimensional. Let $x_1, x_2, \ldots$ be a sequence in $[0, 1]$ that converges to x_0. We shall show that along some subsequence $G(x_k)$ converges to $G(x_0)$. This will imply the continuity of G. Choose a basis $\{f, g\}$ for E with $f(x_0) = 0$ and

$g(x_0) = 1$. From the hypotheses such a choice is easily seen to be possible. Hence

$$S = \{c_\alpha f + d_\alpha g : \alpha \in A\}$$

for some index set A. Since $x_k \to x_0$ and g is continuous, by passing to a subsequence (specifically, by deleting finitely many terms at the beginning of $\{x_k\}$, if necessary} we can assume that $g(x_k) > 0$ for each k. Note that

$$G(x_0) = \sup_\alpha \{d_\alpha\}.$$

If $f(x_k) = 0$ for all k sufficiently large, then for these k we have

$$G(x_k) = G(x_0)g(x_k) \to G(x_0)$$

by the continuity of g, and we are done. If not, by passing to a subsequence and possibly changing the sign of f, we can assume that $f(x_k) > 0$ for each k.

By the definition of G,

$$c_\alpha f(x_1) + d_\alpha g(x_1) \le G(x_1), \quad \text{all } \alpha \text{ in } A.$$

Hence

$$\begin{aligned} c_\alpha f(x_k) + d_\alpha g(x_k) &\le [G(x_1) - d_\alpha g(x_1)]\frac{f(x_k)}{f(x_1)} + d_\alpha g(x_k) \\ &= G(x_1)\frac{f(x_k)}{f(x_1)} + d_\alpha\left[g(x_k) - g(x_1)\frac{f(x_k)}{f(x_1)}\right] \\ &\le G(x_1)\frac{f(x_k)}{f(x_1)} + G(x_0)\left[g(x_k) - g(x_1)\frac{f(x_k)}{f(x_1)}\right], \end{aligned}$$

provided that k is so large that the bracketed quantity, which approaches 1 as $k \to \infty$, is nonnegative. A closer look shows that the right side of this inequality has the form $G(x_0) + \epsilon_k$ where $\epsilon_k \to 0$ as $k \to \infty$, while the supremum of the left side over all pairs (c_α, d_α) is $G(x_k)$. Hence

$$G(x_k) \le G(x_0) + \epsilon_k$$

and

$$\limsup_k G(x_k) \le G(x_0).$$

Since G is the supremum of a set of continuous functions, it is lower semicontinuous, i.e.

$$\liminf_k G(x_k) \ge G(x_0).$$

Hence

$$\lim_k G(x_k) = G(x_0)$$

along a subsequence of our original sequence, and the result follows.

Editorial comment. Adam Riese showed that if we also require

$$G^-(x) = \inf_{g \in S} g(x) > -\infty$$

then both G and G^- are continuous under the weaker assumption that E be finite-dimensional.

Solved also by John A. Frohliger, Reiner Martin (West Germany), Adam Riese, Kenneth Schilling, John Henry Steelman, and the proposer.

A Commutative Diagram Involving Free Groups

6594 [1989, 165]. *Proposed by Stephen M. Gersten, University of Utah, Salt Lake City.*

Let G, A, and B be groups with A and B both free groups. Let $\phi\colon A \to G * B$ be a homomorphism and let $\alpha\colon G * A \to G * B$ be given by $\alpha = (i, \phi)$, where i is the injection $G \to G * B$; that is

$$(i, \phi)\colon g_1 a_1 \dots g_n a_n \mapsto g_1 \phi(a_1) \dots g_n \phi(a_n).$$

Consider a commutative diagram

$$\begin{array}{ccc} G * A & \xrightarrow{\alpha} & G * B \\ \downarrow{\scriptstyle p_A} & & \downarrow{\scriptstyle p_B} \\ A & \xrightarrow[\beta]{} & B \end{array}$$

where the vertical arrows send G to $\{1\}$ and are the identities on A and B, respectively. Prove that if β is injective, then α is also injective.

Solution announcement. The solution will appear as an application in the proposer's paper "Hierarchies of 2-complexes and equations over groups", in preparation. The proposer's proof is elaborate, and involves both geometric methods and towers of branched coverings. The editors published this problem with the hope that one of our readers would find an easier approach that had been overlooked, but no such solution was received.

One and Only One Solution

6596 [1989, 265]. *Proposed by S. D. Chatterji and C. El-Hayek, École Polytechnique Fédérale, Lausanne, Switzerland.*

Prove that the differential equation

$$\frac{dx}{dt} = \frac{1}{t^2 + x^2}$$

has a unique solution defined on $\mathbb{R}$ satisfying the initial condition $x(1) = 1$.

Solution by David V. V. Wend, Department of Mathematics, Montana State University, Bozeman. We first prove this with rather general techniques, and then show how a more specialized method provides more precise information. The only singularity of $x' = x'(t) = (t^2 + x^2)^{-1}$ is at $(0, 0)$, and since $x' > 0$ all solutions are strictly increasing. For such "smooth" initial value problems unique solutions are guaranteed to exist by standard theorems and there are simple comparison theorems. Here the domain of a solution will be $\mathbb{R}$ unless its graph has a vertical

asymptote or has the origin as a limit point. Moreover, if

$$f(t,x) \le g(t,x) \quad \text{and} \quad a \le b \,(a \ge b),$$

then the solution of $x' = f(t,x)$, $x(t_0) = a$, is bounded above (below) to the right (left) of t_0 by the solution of $w' = g(t,w)$, $w(t_0) = b$, on their common domain. We apply this comparison theorem with $f(t,x) = (t^2 + x^2)^{-1}$. For $t \ge 1$ and $x \ge 1$, clearly $f(t,x) \le \frac{1}{2}$ so the increasing solution of $x' = f(t,x)$, $x(1) = 1$ is bounded above by $w(t) = (t+1)/2$, the solution of $w' = \frac{1}{2}$, $w(1) = 1$. Also $f(t,x) \le x^{-2}$ for $x \ne 0$, so for $t \le 1$ the function $x(t)$ is bounded below by $w(t) = (3t-2)^{1/3}$, the solution of $w' = w^{-2}$, $w(1) = 1$. Thus $x(t)$ has no vertical asymptotes and has domain $\mathbb{R}$ unless it approaches 0 as $t \to 0^+$.

For $\frac{1}{2} \le x \le 1$ we have $f(t,x) \ge (t^2+1)^{-1}$ so

$$x(t) \le w_1(t) = \arctan t + c_1,$$

where w_1 solves the corresponding comparison equation with $w_1(1) = 1$ and $c_1 = (4-\pi)/4$. When

$$w_1 = \tfrac{1}{2} = \arctan t_1 + (4-\pi)/4,$$

then

$$t_1 = \tan\left(\frac{\pi}{4} - \frac{1}{2}\right) = 0.2934\ldots$$

and $x(t_1) \le \frac{1}{2} = w_1(t_1)$. For $0 \le x \le \frac{1}{2}$ we have $f(t,x) \ge (t^2 + \frac{1}{4})^{-1}$ so

$$x(t) \le w_2(t) = 2\arctan(2t) + c_2$$

where w_2 solves the comparison equation with $w_2(t_1) = \frac{1}{2}$ and

$$c_2 = \frac{1}{2} - 2\arctan(2t_1) = -0.561\ldots.$$

Now $w_2(t) = 0$ when

$$t = t_0 = \frac{1}{2}\tan\left(-\frac{c_2}{2}\right) = 0.144\ldots.$$

Thus $x(t) = 0$ for some $t \ge t_0 > 0$ and $x(t)$ does not tend to 0 as $t \to 0$. Hence $x(t)$ has domain $\mathbb{R}$ and the result follows. We add that the method of the first part of this solution can be used to show that in general every solution has domain $\mathbb{R}$ except the two unique solutions that approach the origin tangent to the x-axis (their domains are the negative and positive reals respectively).

For an approach tailored to this specific equation, interchange the roles of the dependent and independent variables. This yields the Riccati equation

$$\frac{dt}{dx} = x^2 + t^2,$$

and the usual change of variable $t = -z'/z$ linearizes it to

$$z'' + x^2 z = 0,$$

an equation whose general solution (see [3], pp. 145–146) has the form

$$z = x^{1/2}\left(c_1 J_{1/4}(x^2/2) + c_2 J_{-1/4}(x^2/2)\right).$$

If the solutions of the linear equation are normalized by the condition $z(1) = 1$, then clearly $z(1) = 1$, $z'(1) = -1$, are the appropriate initial conditions for the linear equation ($t(1) = 1$). Thus $z = z(x)$ provides a solution to the Riccati equation on intervals about $x = 1$ where $z(x)$ is nonzero. However, a well-known oscillation theorem [1] shows that $z(x)$ is oscillatory on $\mathbb{R}$. Thus we must deal with numbers x_1 and x_2 (where $x_1 < 1 < x_2$), the zeros of $z(x)$ closest to 1 on the left and right respectively. Since nontrivial solutions of the linear equation can have only simple zeros, $z'(x_1) > 0$ and $z'(x_2) < 0$. Hence $t(x)$ tends to $-\infty$ as $x \to x_1^+$ and to $+\infty$ as $x \to x_2^-$. Thus the range of $t(x)$ is $\mathbb{R}$ and we could say that its inverse (cf. the comment below) $x(t)$ is the desired solution with domain $\mathbb{R}$ *if* we knew that $t(x)$ did not pass through $(0, 0)$. Let $u(x)$ be the solution of the Riccati equation with $u(0) = 0$. Then (see, e.g., [**6**], [**4**] pp. 21–22, C1.14, C2.14 and C2.162(10))

$$u(x) = xJ_{3/4}(x^2/2)/J_{-1/4}(x^2/2)$$

and from tables of Bessel functions [5]

$$u(1) = 0.35023184 < 1 = t(1).$$

Thus $t(x)$ does not pass through the origin and the result follows. It has in fact been proved that all solutions of the original equation have domain $\mathbb{R}$ except for the two solutions that are inverses of $u(x)$ and approach the origin tangent to the x-axis. It can be shown [6] that these two solutions have horizontal asymptotes

$$x = \pm 2.003147359\ldots$$

and that the solution $v(t)$ with $v(1) = 0$ has horizontal asymptotes at

$$x = -2.224\ldots, \quad \text{and} \quad x = 0.96979\ldots.$$

The solution $x(t)$ is sandwiched between these, and must have domain $\mathbb{R}$ since it is monotone. This provides yet another variant of the solution.

Editorial comment. A common approach not involving Bessel functions was to (i) show by standard theory that $dt/dx = t^2 + x^2$ has a unique smooth solution $t = t(x)$ on some interval, (ii) show (e.g. via comparison techniques) that $t(x)$ assumed all real values, and then (iii) invert $t = t(x)$. Various arguments were presented to show that $t(0) \neq 0$, and hence that $dt/dx > 0$ for all x. One of the shortest was Nathaniel Grossman's. Assume $t(0) = 0$. Since t is a nondecreasing function of x and $t(1) = 1$, we have $t'(x) \leq t + x$ for $0 \leq x \leq 1$. Hence,

$$[e^{-x}t(x)]|_0^1 \leq \int_0^1 xe^{-x}\,dx$$

or $3 \leq e$, a contradiction.

The most explicit solution received was that of Jean Anglesio who showed that $x(t)$ was the function inverse to

$$t(x) = x \cdot \frac{J_{3/4}(x^2/2) - kJ_{-3/4}(x^2/2)}{kJ_{1/4}(x^2/2) + J_{-1/4}(x^2/2)},$$

where

$$k = \frac{J_{3/4}(\frac{1}{2}) - J_{-1/4}(\frac{1}{2})}{J_{-3/4}(\frac{1}{2}) + J_{1/4}(\frac{1}{2})} = -0.517050167\ldots.$$

Anglesio also provided the values

$$x(0) = -0.9519678985\ldots,$$
$$\lim_{t\to\infty} x(t) = 1.7294440222\ldots,$$

and

$$\lim_{t\to-\infty} x(t) = -2.127718999\ldots.$$

M. S. Klamkin pointed out that this same differential equation with $x(0) = 0$ appeared as E 1113 [1954, 259; 1954, 714; 1955, 443], and that a closely related problem appeared in the seventh Putnam Competition ([2], pp. 235–236). The Putnam reference was also provided by the proposers. For another example of a differential equation that may be studied in detail either by very general techniques, or by the use of specific special functions, see 6551 [1987, 694; 1989, 657].

REFERENCES

1. W. J. Coles, A simple proof of a well-known oscillation theorem, *Proc. Amer. Math. Soc.*, 19 (1968) p. 507.
2. A. M. Gleason, R. E. Greenwood, and L. M. Kelly, The William Lowell Putnam Mathematical Competition, Problems and Solutions 1938–1964, Math. Assoc. of America, Washington, D.C., 1980.
3. E. Jahnke and F. Emde, Tables of Functions with Formulae and Curves, Dover, N.Y., 1945.
4. E. Kamke, Differentialgleichungen, Lösungsmethoden und Lösungen, Chelsea, New York, 1971.
5. United States National Bureau of Standards, Tables of Bessel Functions of Fractional Order, Columbia University Press, New York, 1948.
6. D. V. V. Wend, The interval of existence of the solution of an initial value problem, Technical Report, Montana State University, Bozeman, 1987.

Solved also by K. F. Andersen (Canada), Jean Anglesio (France), Robert Betts, David Borwein (Canada), Alan E. Berger, F. J. Flanigan, Nathaniel Grossman, Michael Hurley, A. A. Jagers (The Netherlands), Charles Kenney, Murray S. Klamkin (Canada), Thomas McDonald, J. Rausen, Nora S. Thornber, and the proposers. Two incomplete solutions were received.

Too Many Triods

6598 [1989, 366]. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Let K be a closed circular disc in the plane R^2. Suppose that f is a continuous map from K into R^2 which is one-to-one on the interior of K. For each point q in $f(K)$, let $N(q)$ denote the cardinality of the set of all points p in K for which $f(p) = q$.

Prove that $N(q) \geq 3$ for at most countably many points q.

Solution by John Cobb, Mathematics Department, University of Idaho, Moscow. Let C denote the boundary circle of K. Since $f(C)$ and $f(K \setminus C)$ are disjoint, it suffices to consider points of $f(C)$. (Here $A \setminus B$ denotes the set of all points in A but not in B).

For q in $f(C)$, let $H(q)$ denote the convex hull of $f^{-1}(q)$ in K. The $H(q)$'s are pairwise disjoint. For if not, there would be four distinct points a, b, a', b' in cyclic order around C such that

$$f(a) = f(a') \neq f(b) = f(b').$$

If A denotes the straight-line segment in K with endpoints a and a', then $f(a) = f(a')$ implies that $f(A)$ is a simple closed curve, and must separate the plane. Let B denote the straight-line segment in K with endpoints b and b', and let c be the intersection of A and B. Since $f(b) = f(b')$, it follows that $f(B \setminus c)$ is a connected set in the complement of $f(A)$, and hence lies entirely in one complementary domain. But f is locally a homeomorphism at c, and $B \setminus c$ contains points locally on each side of A. Hence $f(B \setminus c)$ contains points in each complementary domain of $f(A)$, which a connected set cannot do. Thus A and B must be disjoint, and so must be any two hulls $H(q)$ and $H(q')$.

If $N(q) \geq 3$, then $H(q)$ contains at least three non-collinear points, and hence contains an open set. If the set of all such q were uncountable, the same would be true of the collection of all the corresponding open sets. But (disjointness!) each such set can be designated uniquely by an ordered pair of rational numbers.

Editorial comment. Several solvers noted that this result follows immediately from Moore's theorem that any family of pairwise disjoint triods (a triod is, roughly, a homeomorph of the sans serif letter "T") in the plane is at most countable. In particular, Mark Meyerson provided a sketch of Moore's proof and the reference R. L. Moore, Concerning triods in the plane and the junction points of plane continua, *Proc. Nat. Acad. Sci.* 14(1928), 85–88. Dov Aharonov and Harold S. Shapiro obtained both the present result and also stronger results with stronger smoothness conditions on f (On the topology of certain simply connected domains, *Technion preprint series* no. MT-465, Haifa; see also Dov Aharonov, Hölder conditions and the topology of simply connected domains, *Canad. Math. Bull.* 26(2) (1983), 189–191).

Solved also by Dov Aharonov (Israel) & Harold S. Shapiro (Sweden), D. W. Finley, Richard P. Jerrard, Ilias G. Kastanas, Marcin E. Kuczma (Poland), Paul MacManus, Mark D. Meyerson, Arlo W. Schurle, and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Lectures on Minimal Surfaces, Vol. 1. By Johannes C. C. Nitsche, Cambridge University Press, New York, 1989, xxv + 563 pp.

ROBERT OSSERMAN
Department of Mathematics, Stanford University, Stanford, CA 94305

In 1978, Springer-Verlag published a book by an unknown author, Arthur L. Besse, with the unpromising title, *Manifolds All of Whose Geodesics are Closed*. To the uninitiated, it might have seemed yet another book in yet another narrow subject, of possible interest to a handful of people. Quite the opposite was true. Besse turned out to be a cousin, once removed, of Bourbaki, with ties to the highest families of French Geometry, and the subject of the book touches not only many parts of geometry, but also other important areas of mathematics.

Three years before the Besse book, Springer published Nitsche's *Vorlesungen Über Minimalflächen*. Here, too, the subject matter seemed terribly constrained. In the decade before the book's appearance the relatively narrow subject of minimal surfaces had begun to expand significantly in various directions—to surfaces in higher-dimensional euclidean spaces through the contributions of Chern, to higher-dimensional hypersurfaces through Bombieri-de Giorgi-Giusti-Miranda, to minimal surfaces in spheres with Calabi, Chern, and Lawson, to general Riemannian manifolds via Jim Simons, and into geometric measure theory with Federer-Fleming and Allard-Almgren. Nitsche resisted all temptation to enlarge the scope of his volume, and focused his attention on classical two-dimensional minimal surfaces in three-dimensional euclidean space. Furthermore, he did so at a length of 775 pages, and to top it off: all in German.

Like *Manifolds All of Whose Geodesics are Closed*, Nitsche's *Vorlesungen Über Minimalflächen* overcame its apparently limited nature by first of all exhibiting a rich theory, with many interesting properties and not a few surprises, and second, by strong links with other parts of mathematics, including topology, partial differential equations, complex function theory, calculus of variations, and measure theory. In addition, as Plateau's experiments make clear, mathematical physics lurks always not far beneath the surface. The classic by C. V. Boys on *Soap Bubbles, Their Colours, and the Forces which Mold Them*, and the 1978 book by Cyril Isenberg, *The Science of Soap Films and Soap Bubbles*, emphasize the physical aspects of the subject.

Since the appearance of Nitsche's book, the field of minimal surfaces has further expanded, if not exploded. In the late seventies, Meeks and Yau demonstrated its importance and value in topology, Schoen and Yau in relativity. Leon Simon and later Schoen managed a synthesis of PDE and geometric measure theory, yielding major new advances. Important contributions were made by Harvey and Lawson, including their fundamental new theory of calibrations. All of this only further accelerated during the eighties when a whole new generation of

geometrical measure theorists—mostly students of Almgren—sprang into action, while a school of young topologists also joined the fray.

But the subject of classical minimal surfaces in $\mathbf{R}^3$ was not lying dormant. Among the noteworthy developments were: first, the global analysis approach to Plateau's problem by Böhme, Tomi, and Tromba; second, the theory of stability, initiated by Barbosa/do Carmo, do Carmo/Peng and Fischer-Colbrie/Schoen, that later flowered into the currently active subject of the index of minimal surfaces, started by Fischer-Colbrie and Gulliver/Lawson; third, the theory of complete *embedded* minimal surfaces in $\mathbf{R}^3$, due initially to Jorge-Meeks, and later blossoming under Costa, Hoffman, Meeks and others; fourth, the breakthroughs on the problem of the Gauss map of complete minimal surfaces, first by Xavier and then Fujimoto.

Meanwhile, behind the scenes, there was Nitsche's book, the source of a wealth of background material for some, and of mainly frustration to others for whom the language was an obstacle. At some point it became clear that a translation would be a worthwhile project. Arrangements between publishers were made, and Jerry Feinberg undertook the literally thankless task of providing an English translation of the German text. However, the many advances we have just described (and others) made the original text terribly out of date in places. Faced with that, and some dissatisfaction with the manner in which many of his carefully formulated and sometimes colloquial or colorful phrases had been turned into English, Nitsche decided to get actively involved once more. He added 12 pages of new Preface, a number of updates inserted in the text, 65 pages of appendices, and over 350 new items to the bibliography. It soon became clear that the expanded version was a bit much for a single volume. What we have so far is the first of two volumes, devoted to the translation of Chapters I–V of the original, and including the additions already mentioned.

By far the bulk of the present volume is devoted to the classical Plateau problem: finding a minimal surface with a given Jordan curve as boundary. Courant's 1950 book, *Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces* (reprinted by Springer in 1977) might easily have given the impression that the last word had been said on the subject. After describing earlier partial results, Courant says: "Complete and satisfactory success was however achieved only in 1930 and 1931, independently, by T. Radó and J. Douglas." Many pages later one finds a reference to three "unsolved problems": boundary behavior, uniqueness, and interior regularity. The third of these revolved around the fact that the solutions of Douglas and Radó were allowed to have certain singularities, called *branch points*, that were in fact never observed in experiments with soap films. How Courant could have considered Plateau's problem "completely and satisfactorily" resolved with that gap still remaining becomes clear if one reads further and realizes that Courant believed that the solution surfaces really did have branch points, so that what remained was not to prove the regularity of the surface, but the (perhaps) less serious problem of finding conditions on the boundary curve for which branch points did or did not occur. In any case, one now knows that the solution surfaces of Douglas and Radó are in fact regularly immersed disks, with no branch points or singularities, and Nitsche includes in the English edition a good deal of new material elucidating the point. There have been

many important contributions to the other two "unresolved problems" mentioned by Courant. However, especially regarding uniqueness, there is much still to be done. In an appendix to the new edition, Nitsche includes a proof of Tomi's theorem that a regular analytic Jordan curve can bound only a finite number of area-minimizing solutions to the Plateau Problem. Whether there can be an infinite number of solutions altogether (including non-area-minimizing ones) is still unknown, even for real analytic boundary curves.

The second volume, promised soon, required even greater revisions and additions to include the newer results in active areas such as the theory of complete minimal surfaces, the subject matter of one of the four remaining chapters. We are also informed of a Volume Three "now in preparation," but with no hint as to its content. In the meantime, Nitsche has given us the latest installment—and a very substantial one—in the never-ending story of minimal surfaces in old-fashioned Euclidean three-space.

I should like to repeat a comment I made in reviewing the original German edition: not only does Nitsche provide us with a great resource in his extensive bibliography, but he enhances enormously the usefulness of the bibliography, and the book as a whole, by indicating after each item listed, all the places in the text where it is mentioned. He is also very good about including precise page references when indicating sources of theorems and proofs. Both of those practices are worthy of emulation, rather than the habit of some of our lazier authors who content themselves with a vague general reference to a 400-page book. The one slightly disconcerting decision Nitsche has made in separating his book into two volumes was to leave references to later sections unamplified, even though the sections referred to may be in a separate (and not yet existent) volume.

But not to end on a peevish note, Nitsche's book has much to offer in a subject area that remains eminently accessible, though highly developed; having it available in an updated and English-language version should assure a wide and enthusiastic audience.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

Education, S(15-16). *How to Use Cooperative Learning in the Mathematics Class.* Alice F. Artzt, Claire M. Newman. NCTM, 1990, v + 73 pp, $6.95 (P). [ISBN: 0-87353-293-7] Intended for high school mathematics teachers, this monograph discusses ways of implementing cooperative learning in the classroom. Sample activities and problem sets are included. SG

Education, P*, L. *Teaching and Learning Mathematics in the 1990s: 1990 Yearbook.* Thomas J. Cooney, Christian R. Hirsch. NCTM, 1990, viii + 216 pp, $18. [ISBN: 0-87353-285-6] The 1990 NCTM yearbook is devoted "to the people of mathematics education—teachers and students—and their changing roles in the face of the calls for reform in the 1990s." The 28 articles are grouped into seven parts: Part 1 is on new perspectives in teaching; Part 2 discusses classroom methods such as small-group work and writing; Part 3 deals with new models and ideas for assessment; Parts 4 and 5 are concerned with cultural and contextual factors in teaching and learning; Part 6 addresses the role of technology in mathematics instruction; and Part 7 discusses the meaning of the word "professionalism" as it relates to mathematics teachers. The authors of the articles represent the entire spectrum of teaching, from elementary through post-secondary. Every teacher of mathematics will certainly find something of interest and importance in this book. MPR

History, L. *Apollonius Conics Books V to VII: The Arabic Translation of the Lost Greek Original in the Version of the Banū Mūsā, Volumes I and II.* Ed. and Trans.: G.J. Toomer. Sources in the Hist. of Math. & Physical Sci., V. 9. Springer-Verlag, 1990, $119 set [ISBN: 0-387-97216-1]. *Volume I: Introduction, Text, and Translation,* xcv + 547 pp, *Volume II: Commentary, Figures, and Indexes,* 337 pp. *Volume I* contains a 95-page introduction and the text in Arabic along with a translation into English. *Volume II* contains commentary, about 300 nicely rendered illustrations, indexes, and a bibliography. JDEK

History, L*. *The Collected Papers of Albert Einstein, Volume 2: The Swiss Years: Writings, 1900-1909.* Eds: John Stachel, *et al.* Princeton University Pr, 1989, xxxii + 656 pp, $85. [ISBN: 0-691-08526-9] Sixty-two papers, reviews, and related documents from Einstein's most productive period, including several of the most influential papers in the history of physics (molecular forces, quantum hypothesis, Brownian motion, relativity, electrodynamics). Extensively annotated with several topical essays provided by the editors—all in English. Einstein's papers are in German; a separate typescript un-annotated translation is under preparation. Second of anticipated 30 volumes. (*Volume 1,* TR, December 1988.) LAS

History, P*, L.** *Joseph Liouville 1809-1882: Master of Pure and Applied Mathematics.* Jesper Lützen. Stud. in History of Math. & Physical Sci., V. 15. Springer-Verlag, 1990, xix + 884 pp, $98. [ISBN: 0-387-97180-7] A massive, thoroughly researched biography of Liouville, in two parts: an account in six chapters of his scientific and professional career—written broadly for non-specialist readers—followed by eleven chapters describing his mathematical contributions to various fields. Numerous quotations (translated into English) from Liouville's extensive unpublished notebooks enliven Lützen's account of the academic and political intrigues of nineteenth century France. Extensive notes link the text to a lengthy bibliography of both published and unpublished references. LAS

Logic, T(14-16). *First Order Mathematical Logic.* Angelo Margaris. Dover, 1990, x + 211 pp, $6.95 (P). [ISBN: 0-486-66269-1] Corrected reprint of the 1967 Blaisdell original (TR, June-July 1968). Two substantive parts: predicate calculus and first order theories. Concludes with decidability and a sketch of a proof of Gödel's theorem. LAS

Logic, P, L. *Reflections on Kurt Gödel.* Hao Wang. MIT Pr, 1988, xxvi + 336 pp, $12.95 (P). [ISBN: 0-262-23127-1] Paperback version of the 1987 original edition (TR, January 1988). A mixture of biography, personal recollections, and intellectual self-expression by a famous logician who was a close professional associate of Gödel. LAS

Graph Theory, S(18), P, L. *Theory of Finite and Infinite Graphs.* Dénes König. Transl: Richard McCoart. Birkhauser Boston, 1990, 426 pp, $75. [ISBN: 0-8176-3389-8] Reprint of the first book on graph theory published in 1936 and containing many classical problems. Results updated in commentary by W.T. Tutte, plus biographical sketch of D. König by T. Gallai. Important for history of field. JPH

Graph Theory, T*(16-17), S, P, L. *Distance in Graphs.* Fred Buckley, Frank Harary. Addison-Wesley, 1990, xiii + 335 pp. [ISBN: 0-201-09591-2] An introductory textbook based on the classic *Graph Theory* by Frank Harary, updated and unified by the concept of distance in graphs. Theorems are clearly stated; proofs are carefully explained and motivated, usually with accompanying figures to elucidate and illustrate concepts. Exercises, large bibliography. LCL

Combinatorics, P. *Coding Theory and Design Theory, Part II: Design Theory.* Dijen Ray-Chaudhuri. IMA, V. 21. Springer-Verlag, 1990, xiii + 378 pp, $39. [ISBN: 0-387-97231-5] Proceedings of a workshop on coding theory and design theory held May 1-June 10, 1988 during the year in applied combinatorics at the IMA, Minneapolis, Minnesota. This volume concentrates on design theory. LC

Discrete Mathematics, S, P, L. *Ramsey Theory, Second Edition.* Ronald L. Graham, Bruce L. Rothschild, Joel H. Spencer. Wiley, 1990, xi + 196 pp, $49.95. [ISBN: 0-471-50046-1] The definitive work on Ramsey Theory, this edition gives a complete treatment of Saharon Shelah's noninductive proof of van der Waerden's Theorem. Also, new material on several more detailed areas, including Graph Ramsey Theory and Euclidean Ramsey Theory. The final chapter relates Ramsey Theory to areas other than discrete mathematics (e.g., the unprovability results of Paris and Harrington). (*First Edition,* TR, March 1981.) LCL

Discrete Mathematics, S, L. *Implementing Discrete Mathematics.* Steven Skiena. Addison-Wesley, 1990, x + 334 pp, $43.25. [ISBN: 0-201-50943-1] Reference on combinatorial algorithms and a lab manual for experimentation in discrete mathematics. Contains over 200 functions written in *Mathematica.* These functions include algorithms for generating permutations, partitions, Young Tableaux , and creating a variety of graph embeddings. Exercises make use of functions described in the book. Useful as supplement for a combinatorics or graph theory course or as an independent study course. (Does not develop the theory behind the functions described.) LC

Number Theory, T(18), S, P***, L**.** *Elementary and Analytic Theory of Algebraic Numbers, Second Edition.* Władysław Narkiewicz. Springer-Verlag, 1990, xii + 746 pp, $89.50. [ISBN: 0-387-51250-0] Updated edition (*First Edition,* TR, June-July 1974) of an important resource. Intended to "reconcile the two approaches" to algebraic number theory (the classical Dedekind ideal theory and local methods). Each chapter ends with exercises and extensive notes for further reading. Appendix includes updated "unsolved problems." Besides updating most chapters to account for recent results, notes at the end of the chapter were rewritten to include literature through 1983, as has the 166(!) page bibliography. Assumes knowledge of topology and algebra (including some Galois theory). A nice reference for expert and nonexpert alike. SB

Number Theory, T(17-18), S, P, L**.** *Divisor Theory.* Harold M. Edwards. Birkhauser Boston, 1990, xiv + 166 pp, $29.50. [ISBN: 0-8176-3448-7] Introduction to Kronecker's Theory of Divisors, a precursor to Dedekind's Ideal Theory (indeed, divisors and ideals coincide in algebraic number fields). Covers the general theory (including some material not treated by Kronecker), and applications to algebraic number theory and algebraic curves (appendix contains proof of Riemann-Roch theorem for curves). Typeset with TEX using large type, making it even more enjoyable to read. No exercises. SB

Number Theory, T(18), S, P, L. *Hilbert Modular Forms.* Eberhard Freitag. Springer-Verlag, 1990, viii + 250 pp, $49. [ISBN: 0-387-50586-5] Provides "a description of the singular cohomology of the Hilbert modular group and its Hodge decomposition including explicit formulae." Each section begins with a brief abstract, adding to its readability. Assumes knowledge of basic complex analysis and algebra; the appendix covers other preliminaries. SB

Number Theory, S(16-18), P, L. *Number Theory and Cryptography.* Ed: J.H. Loxton. London Math. Soc. Lect. Note Ser., V. 154. Cambridge University Pr, 1990, xi + 235 pp, $27.95 (P). [ISBN: 0-521-39877-0] Twenty-two papers on number-theoretic aspects of cryptography, including a lengthy chapter on sieves. BC

Number Theory. *Elementary Theory of Numbers.* William J. LeVeque. Dover, 1990, viii + 132 pp, $4.95 (P). [ISBN: 0-486-66348-5] A reprint of the 1962 edition. Intended as a text for freshmen or sophomore seminars and for teacher refresher courses. Covers Euclidean algorithms, congruences, continued fractions, Gaussian integers, and Diophantine equations. Contains relatively few exercises. SG

Number Theory, T(18), P. *A Classical Introduction to Modern Number Theory, Second Edition.* Kenneth Ireland, Michael Rosen. Grad. Texts in Math., V. 84. Springer-Verlag, 1990, xiv + 389 pp, $49.95. [ISBN: 0-387-97329-X] Adds two chapters to the 1981 edition (TR, January 1983; Extended Review, May 1984), one containing an elementary proof of the Mordell-Weil Theorem, the other chronicling the huge progress made in arithmetic geometry in the '80s. The rest of the book remains essentially the same—a wonderful introduction to elementary number theory, reciprocity laws, zeta functions of varieties, number fields, L-functions, diophantine equations, and elliptic curves. SG

Number Theory, S(17). *Abstract Analytic Number Theory.* John Knopfmacher. Dover, 1990, xi +

336 pp, $9.95 (P). [ISBN: 0-486-66344-2] A reprint with additional bibliography of the 1975 North-Holland edition (TR, May 1976). The focus of the book is on analytic properties of multiplicative semigroups that arise in number theory. SG

Linear Algebra, S(16-18), P*, L. *Matrix Theory and Applications.* Ed: Charles R. Johnson. Proc. of Symposia in Appl. Math., V. 40. AMS, 1990, x + 260 pp, $57. [ISBN: 0-8218-0154-6] Based on the AMS Short Course given at the January 1989 Phoenix meeting. The purpose: to present a sample of the ways in which modern matrix theory is stimulated by its interplay with other subjects. The "other subjects" represented in these seven papers are combinatorics, probability theory, statistics, operator theory and control theory, algebraic coding theory, partial differential equations, and analytic function theory. LCL

Linear Algebra, T(17-18: 2), S, P. *Angewandte Lineare Algebra.* Bertram Huppert. Walter de Gruyter, 1990, viii + 646 pp, DM 198. [ISBN: 3-11-012107-7] Assumes considerable knowledge of linear algebra. Chapters on finite-dimensional Hilbert spaces, linear differential and difference equations, non-negative matrices, geometric algebra, and special relativity. Many problems. JD-B

Linear Algebra, T(15-17: 1), S, L. *Abstract Linear Algebra.* Morton L. Curtis. Universitext. Springer-Verlag, 1990, x + 168 pp, $23 (P). [ISBN: 0-387-97263-3] A rather sophisticated introduction, "directed to students with mathematical leanings," covering the basic results of linear algebra and culminating with the Theorem of Hurwitz (that the only normed algebras over the real numbers are the real numbers, the complex numbers, the quaternions, and the octonions). The emphasis is on the beauty of the mathematics (e.g., noncommutative algebras form the setting for proving the Cayley-Hamilton Theorem, exterior algebras are used in defining determinants), and there are no applications. LCL

Group Theory, T(18: 1), S, P. *Oligomorphic Permutation Groups.* Peter J. Cameron. London Math. Soc. Lect. Note Ser., V. 152. Cambridge University Pr, 1990, viii + 160 pp, $22.95 (P). [ISBN: 0-521-38836-8] Based on notes from a 1988 symposium on Model Theory and Groups; presents recent work on special types of infinite permutation groups. Reference, index. JS

Group Theory, T(18: 1), S, P, L. *Representations and Characters of Finite Groups.* M.J. Collins. Stud. in Adv. Math., V. 22. Cambridge University Pr, 1990, xii + 242 pp, $59.50. [ISBN: 0-521-23440-9] Largely self-contained (except for some block theory); accessible to advanced undergraduates. Two foundation chapters are followed by treatment of Suzuki's theory of exceptional characters, Feit's theory of coherence, Brauer's characterization, and Reynolds and Robinson on isometries. Appendix, exercises, indexes, bibliography. JS

Group Theory, S(18), P. *Induced Modules over Group Algebras.* Gregory Karpilovsky. Math. Stud., V. 161. Elsevier Science, 1990, xi + 520 pp, $120.50. [ISBN: 0-444-88414-9] Intended to present a comprehensive look at the current state of the theory for induced modules, including the substance of a number of recent developments previously available only in research articles. Includes chapters on projective summaries, Green theory, restrictions, permutation modules, and permutation lattices. Bibliography, index. JS

Algebra, S(18), P. *Lecture Notes in Mathematics-1430: Commutative Algebra.* Eds: W. Bruns, A. Simis. Springer-Verlag, 1990, 160 pp, $18 (P). [ISBN: 0-387-52745-1] Eight papers containing the proceedings of the Workshop in Commutative Algebra, Salvador, Brazil, 1988. JS

Algebra, T(15-16: 2), S. *Modern Algebra: Two Volumes Bound as One.* Seth Warner. Dover, 1990, xi + 818 pp, $16.95 (P). [ISBN: 0-486-66341-8] An unabridged, corrected, one-volume version of the original 1965 two-volume work. JS

Algebra, T(18: 1), S, P. *Almost Free Modules: Set-theoretic Methods.* Paul C. Eklof, Alan H. Mekler. Math. Lib., V. 46. North-Holland (US Distr: Elsevier Science), 1990, xvi + 481 pp, $115.50. [ISBN: 0-444-88502-1] A blending of set-theoretic techniques and algebraic methods applied to four problems concerning almost free modules, the structure of Ext, the structure of Hom, and endomorphism rings of groups. Exercises, open problems, bibliography, index. JS

Algebra, T(17: 1), L. *Linear Representations of Groups.* Ernest B. Vinberg. Transl: A. Iacob. Ser. of Adv. Textbooks in Math., V. 2. Birkhauser Boston, 1989, vi + 146 pp, $29.50. [ISBN: 0-8176-2288-8] Fundamentals of linear representation theory, mainly of finite and compact groups, and the elements of Lie groups. Includes a section on the Laplace spherical functions and the representations of SU_2 and SO_3. Contains numerous examples, exercises, and solutions. LC

Algebra, T(15-17: 2), S, L. *Undergraduate Algebra, Second Edition.* Serge Lang. UTM. Springer-Verlag, 1990, xi + 367 pp, $39.80. [ISBN: 0-387-97279-X] This edition includes the Sylow theorems and additional material on symmetric polynomials, principal rings, Jordan normal form, and field theory. There are several new exercises, and more than half a dozen genuine research questions that can be formulated and motivated within the language of this course. (*First Edition,* TR, June-July 1987.) LCL

Algebra, P. *Lecture Notes in Mathematics-1425: Groups of Self-Equivalences and Related Topics.* Ed: R.A. Piccinini. Springer-Verlag, 1990, 214 pp, $21.80 (P). [ISBN: 0-387-52658-7] The proceedings of a conference held at the University of Montreal in 1988. In addition to the research articles, there are two special articles: a survey of groups of self-equivalences, and a paper presenting seventeen open problems. LCL

Algebra, T(18). *Reflection Groups and Coxeter Groups.* James E. Humphreys. Stud. in Adv. Math.,

V. 29. Cambridge University Pr, 1990, xii + 204 pp, $39.50. [ISBN: 0-521-37510-X] Basic theory of Coxeter groups. First part deals concretely with the finite reflection groups and the affine Weyl groups. Includes classification of associated Coxeter graphs, and polynomial invariants of finite reflection groups. Part two develops the general properties of Coxeter groups, including Bruhat ordering. Work of Kazhadan and Lusztig on representations of Hecke algebras associated with Coxeter groups is discussed. LC

Algebra, T(17), L. *Notes on Lie Algebras.* Hans Samelson. Universitext. Springer-Verlag, 1990, xii + 162 pp, $29.80 (P). [ISBN: 0-387-97264-1] Revision of 1969 edition (TR, April 1970). Errors corrected, some text rewritten. New material includes the formulas of Freudenthal and Klimyk for the multiplicities of weights, Brauer's algorithm for the splitting of tensor products, and the Bose-Patera proof of which representations can be brought into orthogonal or symplectic form. LC

Algebra, S(18), P. *Symmetric and G-algebras With Applications to Group Representations.* Gregory Karpilovsky. Math. & Its Applic., V. 60. Kluwer Academic, 1990, xvi + 368 pp, $99. [ISBN: 0-7923-0761-5] Theory of symmetric and G-algebras with applications to modular group representation theory and block theory of group algebras. Essentially self-contained. LC

Algebra, P, L. *Nathan Jacobson: Collected Mathematical Papers, Volumes 1-3.* Nathan Jacobson. Birkhauser Boston, 1989, $225 set. *Volume 1 (1934-1946),* xviii + 454 pp, [ISBN: 0-8176-3410-X]; *Volume 2 (1947-1965),* xviii + 556 pp, [ISBN: 0-8176-3411-8]; *Volume 3 (1965-1988),* xviii + 596 pp. [ISBN: 0-8176-3446-0] Complete collection of Jacobson's papers, both research and expository. Each volume begins with an autobiographical account of the specified period. KS

Algebra, P. *The Orbit Method in Representation Theory.* M. Duflo, N.V. Pedersen, M. Vergne. Progress in Math., V. 82. Birkhauser Boston, 1990, x + 227 pp, $39. [ISBN: 0-8176-3474-6] Ten papers from the proceedings of a conference held at the University of Copenhagen in 1988. The orbit method plays a major role in the representation theory of Lie groups and Lie algebras. LCL

Algebra, S(16-18), P, L. *Numbers.* H.-D. Ebbinghaus, *et al.* Transl. and Ed.: H.L.S. Orde, J.H. Ewing. Grad. Texts in Math., V. 123. Springer-Verlag, 1990, xviii + 393 pp, $59. [ISBN: 0-387-97202-1] A must for any serious student of mathematics. Numbers and number systems—natural, integral, real, complex, p-adic, quaternion, Cayley, nonstandard, and Conway—are described in detail. Many major theorems (the Fundamental Theorem of Algebra, the Gelfand-Mazur Theorem, the transcendence of pi) are proved, and historical notes are sprinkled thickly throughout. SG

Algebra, T*(15-16), S*, L*. *Galois Theory.* Joseph Rotman. Universitext. Springer-Verlag, 1990, xii + 108 pp, $29 (P). [ISBN: 0-387-97305-2] A considered and cohesive introduction to the main results of Galois theory: the cubic and quartic formulas, the fundamental theorem, the insolvability of the quintic, and the "Great Theorem" that solvability of a polynomial by radicals is equivalent to the solvability of its Galois group. The exposition builds on an introductory knowledge of linear and abstract algebra, leaving advanced topics in field theory out altogether if not directly needed. Especially suitable for a senior or first-year graduate seminar or for self-study. LCL

Calculus, T(13), L. *Precalculus: Functions and Graphs, Second Edition.* Bernard Kolman, Arnold Shapiro. Harcourt Brace Jovanovich, 1990, xvi + 645 pp, $31. [ISBN: 0-15-571063-X] This edition adds chapters on maxima and minima and root finding. (*First Edition,* TR, January 1985.) OJ

Calculus, P, L. *Priming the Calculus Pump: Innovations and Resources.* Ed: Thomas W. Tucker. MAA Notes No. 17. MAA, 1990, 321 pp, $20 (P). [ISBN: 0-88385-067-2] In-depth report on twelve innovative calculus projects—complete with samples of problem sets, tests, and student work—together with extensive collateral information (over 60 abstracts of other projects, list of NSF awards, database of software for college mathematics) on the calculus reform movement. Third in MAA's calculus reform series, following *Towards a Lean and Lively Calculus* (1986) and *Calculus for a New Century* (1987), this volume provides a wealth of detail from real classrooms with real students. You will be amazed at what calculus students can accomplish with the right tools. LAS

Calculus, S*(13), P. *Calculus Activities for Graphic Calculators.* Dennis Pence. PWS-Kent, 1990, x + 266 pp, $15 (P). [ISBN: 0-534-92431-X] Clear and concise instructions and examples (including display output) for Casio, Sharp, and HP graphic scientific calculators; Chapters 1-4 introduce basic calculator operations and graphing; remaining chapters illustrate graphical and numerical approaches to first-year calculus topics. Exercises. JNC

Real Analysis, T(16-17: 1, 2), L. *Analysis on Manifolds.* James R. Munkres. Addison-Wesley, 1991, xi + 366 pp, $48.50. [ISBN: 0-201-51035-9] Orderly, clearly written presentation of the theory of derivatives and the Riemann integral for multivariable functions, and of the theory of integration of forms on manifolds in Euclidean space. Like the author's other texts, this book should be a good one from which to learn a body of standard material. JO

Complex Analysis, T(15-16), L. *Complex Variables for Scientists and Engineers, Second Edition.* John D. Paliouras, Douglas S. Meadows. Macmillan, 1990, xiii + 586 pp. [ISBN: 0-02-390561-1] Fairly standard, applications-oriented text appropriate for a first course aimed at engineering majors. Many examples illustrating theorems and many exercises of varying degrees of difficulty. One complaint: it takes over 200 pages to get to the Cauchy integral

formula, and almost 250 pages to get to series. (*First Edition,* TR, November 1975.) MPR

Complex Analysis, P. *Lecture Notes in Mathematics-1422: Complex Geometry and Analysis.* Ed: V. Villani. Springer-Verlag, 1990, v + 109 pp, $14.70 (P). [ISBN: 0-387-52434-7] Proceedings from a symposium held in Pisa during May 1988. Ten papers intended to survey current research in complex geometry, including essays on hyperkähler and *CR*-manifolds and complex vector bundles. JO

Partial Differential Equations, P. *Lattice Gas Methods for Partial Differential Equations.* Eds: Gary D. Doolen, *et al.* Addison-Wesley, 1990, xix + 554 pp, $47.50; $26.95 (P). [ISBN: 0-201-15679-2] Twenty-nine papers on lattice gas methods, an up-and-coming numerical approach to solving partial differential equations, especially for hydrodynamics. BC

Partial Differential Equations, P. *Solitons in Physics, Mathematics, and Nonlinear Optics.* Eds: Peter J. Olver, David H. Sattinger. IMA, V. 25. Springer-Verlag, 1990, xiii + 215 pp, $29. [ISBN: 0-387-97309-5] Solitons, or solitary waves, are solutions of certain partial differential equations that have many surprising, particle-like properties. This book contains eleven papers from a 1988 workshop at the Institute for Mathematics and Its Applications. BC

Numerical Analysis, T(16-17: 1), P. *Éléments D'Analyse Numérique.* Marc Atteia, Michel Pradel. Cepadues-Editions, 1990, viii + 154 pp, 90F (P). [ISBN: 285428-229-9] Concise, demanding introductory text. Treats polynomial approximation and interpolation, initial-value problems, linear systems. Problems. JD-B

Numerical Analysis, T(15). *Numerical Analysis, An Introduction.* Lars Eldén, Linde Wittmeyer-Koch. Academic Pr, 1990, x + 347 pp, $39.95. [ISBN: 0-12-236430-9] Intended as a text for a standard undergraduate course, this is a revision of a book originally published in Swedish. It is written from the viewpoint that students will best understand algorithms if they have worked through them using a hand-held calculator, and exercises and examples are constructed accordingly. AWR

Functional Analysis, T(17-18: 2, 3), S, L. *A Course in Functional Analysis, Second Edition.* John B. Conway. Grad. Texts in Math., V. 96. Springer-Verlag, 1990, xvi + 399 pp, $49.50. [ISBN: 0-387-97245-5] Most significant change in this edition is that the chapter on Fredholm Theory has been completely rewritten and simplified. Bibliographical notes and (more) exercises have been added. (*First Edition,* TR, December 1985; Extended Review, November 1986.) KS

Functional Analysis, T(18: 2), P. *Volterra Integral and Functional Equations.* G. Gripenberg, S.-O. Londen, O. Staffans. Encyclop. of Math. & Its Applic., V. 34. Cambridge University Pr, 1990, xxii + 701 pp, $99.50. [ISBN: 0-521-37289-5] A new book intended as a text (there are problem sets) and as a book for researchers in the field (there are extensive references, and an effort has been made to bring the reader up to current research). Efforts have been made to cover aspects of analysis needed but not easily found elsewhere (the reader is assumed to be familiar with the contents of, say, Rudin's books), and care is given to including an extensive table of symbols and their meaning; it is nicely produced. The three parts cover linear theory, quasi-linear and existence problems for nonlinear equations, and frequency domain methods in nonlinear equations. AWR

Functional Analysis, P. *Lecture Notes in Mathematics-1421: A Nonlinear Theory of Generalized Functions.* Hebe A. Biagioni. Springer-Verlag, 1990, xii + 214 pp, $21.80 (P). [ISBN: 0-387-52408-8] Presents a theory (due to J.F. Colombeau) for incorporating distributions into a differential algebra, which Laurent Schwartz proved impossible in 1954. (Schwartz wasn't wrong; Colombeau weakens certain properties.) Includes applications to numerical methods in fluid dynamics, elastoplasticity, and acoustics. BC

Functional Analysis, P. *Uniform Frechet Algebras.* Helmut Goldmann. Math. Stud., V. 162. Elsevier Science, 1990, viii + 355 pp, $102.50. [ISBN: 0-444-88488-2] A study of algebras of complex-valued continuous functions on hemicompact spaces which are complete in the compact open topology. Compares with the theory of uniform (Banach) algebras. MLR

Functional Analysis, T(16-17: 1-3), L. *Introduction to Hilbert Spaces with Applications.* Lokenath Debnath, Piotr Mikusiński. Academic Pr, 1990, xiv + 509 pp, $49.95. [ISBN: 0-12-208435-7] First half of this friendly text contains basics of Hilbert space theory and of linear operators on Hilbert spaces, including simple introduction (without concept of measure) to Lebesgue integration. Later, Hilbert space formalism is used to develop foundations of quantum mechanics, and many applications of Hilbert space methods are given. Contains wide selection of examples and exercises. KS

Functional Analysis, T(16-17: 2, 3), S, L. *Functional Analysis.* Frigyes Riesz, Béla Sz.-Nagy. Transl: Leo F. Boron. Dover, 1990, xii + 504 pp, $11.95 (P). [ISBN: 0-486-66289-6] Paperback reprint of classic text. Part One takes up differentiation, the Lebesgue integral, and the Stieltjes integral and its generalizations. Part Two addresses integral equations, Hilbert and Banach spaces, transformations of Hilbert space, self-adjoint transformations, groups and semigroups of transformations, and spectral theories for linear transformations of general type. KS

Analysis, P, L. *Basic Hypergeometric Series.* George Gasper, Mizan Rahman. Encyclop. of Math. & Its Applic., V. 35. Cambridge University Pr, 1990, xx + 287 pp, $59.50. [ISBN: 0-521-35049-2] Valuable text and reference book on the significant results of basic hypergeometric series from the past

200 years. Chapter topics include introduction of definitions and notations of hypergeometric and basic hypergeometric series and the early work of Heine, Jackson, and Bailey; summation, transformation and expansion formulas; basic contour integrals; bibasic series; the Askey-Wilson q-beta integral and applications to orthogonal polynomials. Prerequisites are a good background in analysis. Exercises provide additional formulas and practice with proof techniques. LC

Analysis, T(17-18: 2), P, L. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory, Second Edition.* Theodore J. Rivlin. Wiley, 1990, xiii + 249 pp, $49.95. [ISBN: 0-471-62896-4] Aims to give a survey of important properties of the Chebyshev polynomials and introduce some interesting areas of mathematics (e.g., interpolation theory, numerical integration, ergodic theory) where the Chebyshev polynomials serve as an example. Revisions include new applications, a new chapter on algebraic and number theoretic properties of the Chebyshev polynomials, and new exercises (*First Edition,* TR, May 1975). LC

Analysis, P. *Asymptotic and Computational Analysis: Conference in Honor of Frank W.J. Olver's 65th Birthday.* Ed: R. Wong. Lect. Notes in Pure & Appl. Math., V. 124. Marcel Dekker, 1990, xii + 755 pp, $115 (P). [ISBN: 0-8247-8347-6] Proceedings of the International Symposium on Asymptotic and Computational Analysis held June 1989 in Winnipeg, Canada. LC

Analysis, T(17-18), P. *Unitary Representations and Harmonic Analysis: An Introduction, Second Edition.* Mitsuo Sugiura. Math. Lib., V. 44. Elsevier Science, 1990, xv + 452 pp, $112.75. [ISBN: 0-444-88593-5] Introduction to harmonic analysis and theory of unitary representations of (Type I) Lie groups. This edition includes a section on irreducible unitary representations of $SL(2, R)$. Other changes include a handful of rewritten proofs and a bibliography tripled in length. (*First Edition,* TR, May 1976.) KS

Analysis, P, L. *Iterative Functional Equations.* Marek Kuczma, Bogdan Choczewski, Roman Ger. Encyclop. of Math. & Its Applic., V. 32. Cambridge University Pr, 1990, xix + 552 pp, $99.50. [ISBN: 0-521-35561-3] Topics covered include systems of linear and nonlinear equations of finite and infinite order in various function classes, conjugate and commutable functions, linearization, iterative roots of functions, and special functional equations. Many results have appeared before only in research literature. Contains an extensive bibliography. KS

Analysis, S(18), P. *Quasiconformal Mappings and Sobolev Spaces.* V.M. Gol'dshtein, Yu. G. Reshetnyak. Math. & Its Applic., V. 54. Kluwer Academic, 1990, xix + 371 pp, $133. [ISBN: 0-7923-0543-4] A relatively self-contained introduction to basic methods and recent results on Sobolev spaces, various geometric classes of mappings (including quasiconformal and quasiisometric mappings), and nonlinear capacity. The first three chapters comprise a readable (occasional quirks of translation aside), inviting, and fairly leisurely introduction to modern analysis: integration, differentiation, distribution theory, Sobolev spaces, and other function spaces. Note price. PZ

Analysis, S, P, L*. *Problems in Mathematical Analysis.* Piotr Biler, Alfred Witkowski. Pure & Appl. Math., V. 132. Marcel Dekker, 1990, v + 227 pp, $49.75. [ISBN: 0-8247-8312-3] More than 1300 problems in analysis, most of them neither standard nor easy (at the level of the *Monthly* Problems Section and beyond), covering the usual topics in analysis and including measure and integration, functional analysis, Fourier analysis, functions of several variables, analytic functions. The answers and hints are extremely laconic; many are references to the problems literature (especially journals). LCL

Analysis, S(17-18), P. *Operator Theory, Operator Algebras and Applications.* Eds: William B. Arveson, Ronald G. Douglas. Proc. of Sympos. in Pure Math., V. 51. AMS, 1990, $174 set [ISBN: 0-8218-1486-9]. *Part 1,* xii + 640 pp; *Part 2,* xii + 385 pp. Proceedings of a 1988 Summer Institute held in Durham, New Hampshire. KS

Algebraic Geometry, P. *Lectures on the Mordell-Weil Theorem.* Jean-Pierre Serre. Transl: Martin Brown, Michel Waldschmidt. Aspects of Math., V. E15. Friedr Vieweg (US Distr: GLP Intern. Vieweg), 1990, x + 218 pp. [ISBN: 3-528-18968-1] Notes from a pre-Faltings course. Topics include the Mordell-Weil Theorem, Mordell's conjecture, integral points on curves via Siegel's method and Baker's method, Hilbert's irreducibility theorem, construction of Galois extension, and large sieves. SG

Differential Geometry, P. *Symplectic Geometry.* A.T. Fomenko. Trans: R.S. Wadhwa. Adv. Stud. in Contemp. Math., V. 5. Gordon & Breach, 1988, xiii + 387 pp, $180. [ISBN: 2-88124-657-5] Basic theory and some recent results related to the structure of symplectic manifolds, the embedding of symplectic and Lagrangian manifolds, and the complete integrability of Hamiltonian systems. Too compact to make a good textbook, but contains some nice reference sections, including tight over views of matrix groups, symplectic structures on manifolds, and Hamiltonian systems with symmetries. Note the price. JO

Differential Geometry, T(18: 2), S, P, L. *Lie Groups and Algebraic Groups.* A.L. Onishchik, E.B. Vinberg. Transl: D.A. Leites. Springer-Verlag, 1990, xix + 328 pp, $79. [ISBN: 0-387-50614-4] "The only book where the theory of semisimple Lie groups is based systematically on the technique of algebraic groups." Theory is developed as a sequence of problems (with hints), leading eventually to classification of real and complex semisimple Lie groups and Levi-Malcev theorems for Lie and algebraic groups. Tables, exercises, bibliography, index. JS

Geometry, T, S*(13-15), P, L*. *Connections: The Geometric Bridge between Art and Science.* Jay Kappraff. McGraw-Hill, 1991, xxi + 471 pp, $19.50 (P); $39.50. [ISBN: 0-07-034251-2; 0-07-034250-4]

A spectacular presentation of design science—"the grammar of space"—that explores with rich details instances of similarity, proportion, tilings, graphs, lattices, polyhedra, isometries, and symmetry in art, architecture, engineering, and science. Punctuated with exercises and problems (thus making the monograph useful as a course or seminar text); illustrated with over 200 figures; supported by an extensive multi-disciplinary bibliography that is well-referenced to the text. A superb option for interdisciplinary seminars. LAS

Topology, P. *General Topology and Applications: Proceedings of the 1988 Northeast Conference.* Ed: R.M. Shortt. Lect. Notes in Pure & Appl. Math., V. 123. Marcel Dekker, 1990, xiv + 291 pp, $99.75 (P). [ISBN: 0-8247-8349-2] More than twenty papers covering a wide range of research topics representing the topological interests of Melvin Henriksen, for whom this conference was organized, on the occasion of his sixtieth birthday. LCL

Dynamical Systems, P, L*. *Dynamical Chaos.* Eds: M.V. Berry, I.C. Percival, N.O. Weiss. Princeton University Pr, 1987, 199 pp, $19.50 (P). [ISBN: 0-691-02423-5] Proceedings of a 1987 discussion meeting of the Royal Society of London, reprinted from their *Proceedings*. Introduced by E.C. Zeeman; contains survey papers with various applications, including cardiac arrhythmias, biological populations, convection, solar system, Hamiltonian systems, and renormalization. LAS

Dynamical Systems, T(18: 2), L. *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* Stephen Wiggins. Texts in Appl. Math., V. 2. Springer-Verlag, 1990, xiv + 672 pp, $49.95. [ISBN: 0-387-97003-7] Intended as a graduate text with sufficient completeness to prepare students for research in the field. An informal expositional style and constant attention to the development of the reader's intuition make this a most readable text. JO

Dynamical Systems, S(17-18), P. *Hamiltonian Systems: Chaos and Quantization.* Alfredo M. Ozorio de Almeida. Mono. on Math. Physics. Cambridge University Pr, 1990, ix + 238 pp, $24.95 (P). [ISBN: 0-521-34531-6] A physics-oriented introduction to chaos in Hamiltonian systems, focusing on neighborhoods of periodic orbits, and implications for quantum dynamics. (1988 hardcover text, TR, November 1989.) BC

Dynamical Systems, P. *Geometry in the Neighborhood of Invariant Manifolds of Maps and Flows and Linearization.* U. Kirchgraber, K.J. Palmer. Pitman Res. Notes in Math. Ser., V. 233. Longman Scientific & Technical (US Distr: Wiley), 1990, 89 pp, $32 (P). [ISBN: 0-470-21657-3] Presents detailed proofs of generalized Grobman-Hartman linearization theorems for discrete and continuous dynamical systems. Studies the geometric properties of invariant manifolds and foliations. OJ

Control Theory, P. *Signal Processing, Part II: Control Theory and Applications.* Ed: F.A. Grünbaum, J.W. Helton, P. Khargonekar. IMA, V. 23. Springer-Verlag, 1990, xiii + 411 pp, $39. [ISBN: 0-387-97230-7] Twenty papers from a program at the Institute for Mathematics and Its Applications. Applications include tomography, x-ray problems, and speech recognition. BC

Control Theory, P. *Lecture Notes in Control and Information Sciences-140: Singularly Perturbed and Weakly Coupled Linear Control Systems: A Recursive Approach.* Z. Gajić, D. Petkovski, X. Shen. Springer-Verlag, 1990, vii + 202 pp, $32.90 (P). [ISBN: 0-387-52333-2] Singularly perturbed and weakly coupled linear systems of ordinary differential equations are traditionally studied via non-recursive, and hence computationally expensive, power series methods. Here the authors develop recursive algorithms for solving the associated Lyapunov-Ricatti equations. Most chapters include applications to real-world case studies. PZ

Probability, T(15-17: 2), L. *The Theory of Probability and the Elements of Statistics.* B.V. Gnedenko. Transl: B.D. Seckler. Chelsea, 1989, 529 pp, $23.95. [ISBN: 0-8284-1132-9] *Fifth Edition* of a classic, mathematical probability text; includes a new chapter on mathematical statistics with empirical distribution functions and Glivenko-Cantelli, Kolmogorov-Smirnov theorems. Testing and estimation are dealt with cursorily. Remainder of book covers advanced concepts in probability: CLTs, LLNs, characteristic functions, infinite divisibility, stochastic processes, etc. No measure theory is required, although some is discussed. (1967 Chelsea edition, TR, February 1968; 1976 MIR edition, TR, August-September 1976; 1978 MIR edition, TR, November 1981.) MK

Stochastic Processes, T(17: 1), P, L. *Mathematical Methods for Construction of Queueing Models.* V.V. Kalashnikov, S.T. Rachev. Oper. Res. Ser. Wadsworth, 1990, xvi + 431 pp, $65.95. [ISBN: 0-534-132545] This work is concerned with modelling extant queueing systems. While some consideration is given to the theory of the usual well-known queueing models, the emphasis here is on matching distribution parameters to empirical data, choosing a class of model appropriate to the real-life system, and determining the effects of the inevitable mathematical simplifications and approximations found in any mathematical model. No exercises. SM

Stochastic Processes, T(18). *Stochastic Processes with Applications.* Rabi N. Bhattacharya, Edward C. Waymire. Ser. in Prob. & Math. Stat. Wiley, 1990, xiii + 672 pp, $69.96. [ISBN: 0-471-84272-9] Textbook designed for science and engineering (and mathematics) graduate students. Presentation is technically simple with emphasis on computation and examples. Applications from physics, computer science, economics, and engineering. MLR

Stochastic Processes, P. *Stochastic Equations and Differential Geometry.* Ya. I. Belopolskaya, Yu. L. Dalecky. Math. & Its Applic., V. 30. Kluwer Academic, 1990, xv + 260 pp, $99. [ISBN: 90-277-2807-0] Presents theory of stochastic equations in Ba-

nach space and on manifolds, leading to discussions of Kolmogorov equations and diffusion processes on Lie groups and principal fibre bundles. JO

Elementary Statistics, T(13-16: 1), S, L*. *Statistics for Lawyers.* Michael O. Finkelstein, Bruce Levin. Texts in Stat. Springer-Verlag, 1990, xxii + 608 pp, $39 (P). [ISBN: 0-387-97136-X] A series of several dozen case studies (e.g., Dalkon shield, automobile emissions, death penalty, draft lotteries) arranged by relevant statistical methodology (e.g., descriptive statistics, counting, statistical inference, sampling, regression), interspersed with brief synopses of the relevant theoretical issues and commentary on actual court decisions. Numerous exercises require both calculation (full solutions at the end of the volume) and interpretation. Of interest both to future lawyers and to future statisticians; excellent source of "relevant" examples. LAS

Computational Statistics, P, L. *Advances in Statistical Methods for Genetic Improvement of Livestock.* Eds: Daniel Gianola, Keith Hammond. Adv. Ser. in Agric. Sci., V. 18. Springer-Verlag, 1990, xx + 534 pp, $79. [ISBN: 0-387-50809-0] Edited version of an international symposium which took place in Armidale, Australia, February 16-20, 1987 concerning recent developments in statistics and computing and their application to identifying genetically superior livestock and estimating genetic trends in breeding programs. Methods range from likelihood-inference, Bayesian and empirical Bayesian techniques for nonlinear models to generalized linear models and design of experiments and breeding programs. MK

Statistics, T(16-17: 1, 2), C, L. *Robust Estimation and Testing.* Robert G. Staudte, Simon J. Sheather. Ser. in Prob. & Math. Stat. Wiley, 1990, xix + 351 pp, $54.95. [ISBN: 0-471-85547-2] Well-motivated, practical presentation of robust estimation and testing procedures illustrated with several small data sets. Examines the robustness of these procedures primarily with respect to outliers and fat-tailed distributions. Solid treatment of robust methods of inference in regression. Includes disk containing data sets and Minitab macros to calculate robust estimates and their standard errors. RWJ

Statistics, P. *Spatial Statistics: Past, Present, and Future.* Ed: Daniel A. Griffith. Mono., No. 12. Institute of Math Geography (IMaGe, 2790 Briarcliff, Ann Arbor, MI 48105), 1990, xvi + 398 pp, (P). [ISBN: 1-877751-42-1] Contains spring 1989 Syracuse University symposium proceedings. The thirteen papers summarize recent developments in point and regional processes, and explore various modelling, estimation, testing, and data analysis questions. RWJ

Statistics, P. *Statistical Design and Analysis of Industrial Experiments.* Ed: Subir Ghosh. Stat.: Textbooks & Mono., V. 109. Marcel Dekker, 1990, xi + 533 pp, $55. [ISBN: 0-8247-8251-8] Contains eighteen papers on methods of experimental design and quality control, with some emphasis on factorial designs. Applications papers include analysis of health studies data. RWJ

Statistics, T(17-18: 1, 2), S, P, L. *Regression Estimators: A Comparative Study.* Marvin H.J. Gruber. Statistical Modeling & Decision Sci. Academic Pr, 1990, xi + 347 pp, $49.95. [ISBN: 0-12-304752-8] Derives and examines the relation, efficiencies, and robustness of least squares, ridge, minimax, and Bayes estimates of the parameters of a linear regression model. Applications to the Kalman filter, and one- and two-way ANOVA. RWJ

Statistics, P, L. *A Statistical Model: Frederick Mosteller's Contributions to Statistics, Science, and Public Policy.* Eds: S.E. Fienberg, *et al.* Ser. in Stat. Springer-Verlag, 1990, xviii + 283 pp, $39. [ISBN: 0-387-97223-4] Belatedly honors Mosteller on his seventieth birthday. Includes a brief biography by J.W. Tukey, comprehensive bibliography of Mosteller's books, papers, etc., and reviews of his books. Friends, collaborators, colleagues, and students contribute discussions of Mosteller's contributions to mathematical statistics, scientific generalism, methodology and applications, as well as bibliographical anecdotes on Mosteller as an educator and his years at Harvard. Includes numerous photographs. MK

Statistics, P, L. *Contributions to Econometric Theory and Application: Essays in Honour of A.L. Nagar.* Eds: R.A.L. Carter, J. Dutta, A. Ullah. Springer-Verlag, 1990, xiii + 366 pp, $49. [ISBN: 0-387-97285-4] Collection of papers in honor of A.L. Nagar's sixtieth birthday. Topics range from the role of exogenous variables in structural models, to measurement of the inequality in the distribution of income in developing countries, to dynamic, simultaneous, linear equation models and resampling schemes in linear regression and errors-in-variables models. MK

Statistics, S(18), P, L. *Nonlinear Estimation.* Gavin J.S. Ross. Ser. in Stat. Springer-Verlag, 1990, viii + 189 pp, $39. [ISBN: 0-387-97278-1] Advanced text concerns formulating, fitting, and interpreting non-standard, non-linear models. Mathematical exposition terse, but covers broad spectrum; inference, stable transformations, geometry of non-linear inference, etc. One chapter treats computational methods for non-linear modelling, including computer programs, libraries, and algorithms for non-linear optimization; another contains brief explanations of applications. Many references make for a nice, encyclopedic companion. MK

Statistics, T(17-18: 2, 3), L. *Statistical Inference.* George Casella, Roger L. Berger. Stat. & Prob. Ser. Wadsworth, 1990, xviii + 650 pp, $54.95. [ISBN: 0-534-11958-1] Develops theoretical statistics from first principles of probability theory. Discusses how to evaluate, as well as find, the appropriate statistical techniques for estimation and hypothesis testing problems. Presents use of sufficiency, likelihood, and invariance principles in chapter on data reduction. Other topics: ANOVA, decision theory and

Stein estimation, and linear regression. Nearly 600 exercises. RWJ

Statistics, P*. *The Collected Works of John W. Tukey, Volume VI: More Mathematical: 1938-1984.* Ed: Colin L. Mallows. Stat. & Prob. Ser. Wadsworth, 1990, lxxii + 661 pp, $59.95. [ISBN: 0-534-05103-0] Sixth volume of a series containing Tukey's many and diverse works. Previous volumes covered his contributions in the areas of time series (TR's, May 1985 and February 1986), data analysis (TR, June-July 1987), and graphics (TR, December 1988). This volume contains "six early mathematical works, four papers on fiducial inference, five on transformations, and twenty-seven on a miscellany of topics in mathematical statistics," again with comments by the editor on each of the papers. RSK

Computer Literacy, T, S*(13-16), L. *Computer Ethics: Cautionary Tales and Ethical Dilemmas in Computing.* Tom Forester, Perry Morrison. MIT Pr, 1990, viii + 193 pp, $19.95. [ISBN: 0-262-06131-7] An unrelenting tale of mistakes, mismanagement, and malfeasance in the production and use of computer systems, laced with thought-provoking questions and—at each chapter's end—hypothetical scenarios intended to provoke role-playing debate among students. Covers computer crime, software theft, hacking, reliability, privacy, artificial intelligence, and workplace issues. Raises many ethical questions, but does not examine them from a formal philosophical perspective. Intended merely to sensitize computer science students to ethical issues, not to lay a foundation for a new branch of ethics. LAS

Elementary Computer Science, T(13: 1). *Problem Solving and Structured Programming in FORTRAN 77, Fourth Edition.* Elliot B. Koffman, Frank L. Friedman. Addison-Wesley, 1990, xvi + 652 pp, (P). [ISBN: 0-201-51216-5] A text for a first course in computer programming. High school algebra presents a sufficient mathematical preparation. Contains a large number of applications from business, engineering, and the physical sciences. Nonstandard topics include elementary computer-aided design and numerical methods. Exercises and short programming projects appear at the end of each chapter, along with examples of common programming errors. (*First Edition,* TR, June-July 1981; *Second Edition,* TR, January 1986.) SM

Programming, T(14), S. *An Introduction to Object-Oriented Programming and C++.* Richard S. Wiener, Lewis J. Pinson. Addison-Wesley, 1988, xii + 273 pp, (P). [ISBN: 0-201-15413-7] C++ is an object-oriented extension of the system programming language C. The extensions allow a user to define a hierarchy of classes, and to create objects with the properties of inheritance and polymorphism. This text is an introduction to the principles of object-oriented software development, and the syntax and semantics of the C++ programming language. GMS

Programming, S(13), L. *The C++ Programming Language.* Bjarne Stroustrup. Addison-Wesley, 1987, viii + 328 pp, (P). [ISBN: 0-201-12078-X] Readable introduction to C++ with careful attention to the needs of programmers already fluent in C. Although it is a complete reference, this book concentrates most on those features and data types which are not part of C. JO

Programming, S. *Advanced Prolog: Techniques and Examples.* Peter Ross. Intern. Ser. in Logic Programming. Addison-Wesley, 1989, x + 294 pp. [ISBN: 0-201-17527-4] Prolog is an acronym for *Pro*gramming with *log*ic, and it is a language based on the principles of the first-order predicate logic. This text assumes that the reader has had a brief introduction to the language but is not a sophisticated user. It shows how to use the advanced features of the language to design and build complex software systems. GMS

Programming, T(13-14: 1), L. *Program Design with Modula-2.* Susan Eisenbach, Chris Sadler. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xiii + 285 pp, (P). [ISBN: 0-201-17567-3] Based on courses given at the Polytechnic of North London. Introduction to programming in Modula-2 with emphasis on program design rather than language syntax. LC

Languages, T(16-18: 1), S, L. *Logic Programming: Systematic Program Development.* Yves Deville. Intern. Ser. in Logic Programming. Addison-Wesley, 1990, xvii + 338 pp. [ISBN: 0-201-17576-2] Text is divided into three parts which reflect the problem solving cycle of a Prolog programmer: elaborate a specification for the problem; construct a logic description of the problem; derive a correct and efficient Prolog program. Assumes familiarity with Prolog. Chapter exercises, background sections, and summaries. Appendices; bibliography. RJA

Computer Systems, S(18), P. *Computers in Mathematics.* Eds: David V. Chudnovsky, Richard D. Jenks. Lect. Notes in Pure & Appl. Math., V. 125. Marcel Dekker, 1990, vii + 410 pp, $89.75 (P). [ISBN: 0-8247-8341-7] Fifteen papers from the July 1986 International Conference on Computers and Mathematics held at Stanford. Topics include something for everyone: papers range from modern applications of computer algebra (e.g., Grobner basis methods for solving systems of polynomial equations) to history of computation (George Dantzig on linear programming and its effects on computing) to applications of computing to mathematical physics. PZ

Computer Systems, T(13-14: 1). *Calculator Enhancement for Single-Variable Calculus: A Manual of Applications Using the HP-28S Calculator, Preliminary Edition.* James H. Nicholson. Consultant: J.W. Kenelly. Harcourt Brace Jovanovich, 1991, ix + 99 pp, (P), [ISBN: 0-15-505676-X]; *Calculator Enhancement for Linear Algebra.* D.R. LaTorre, 1991, ix + 130 pp, (P), [ISBN: 0-15-505674-3]; *Calculator Enhancement for Multivariable Calculus.* J.A. Reneke. Ed: D.R. LaTorre. 1991, vii + 94 pp, (P), [ISBN: 0-15-505675-1]; *Calculator Enhancement for Differential Equations.* T.G. Proctor. Ed: D.R.

LaTorre. 1991, xiii + 105 pp, (P). [ISBN: 0-15-505673-5] Four slim volumes on using the HP-28S (and, soon, the HP-48S) calculator for mathematical work in calculus, multivariable calculus, linear algebra, and differential equations. In addition to ideas and exercises on a variety of topics in each course area, each volume contains many examples of useful programs. Taken together they considerably extend the power and ease of use of the calculator for an impressive variety of mathematical operations. Students who use these supplements diligently will indeed "enhance" their understanding both of the calculator and of the underlying mathematics. PZ

Computer Graphics, T(16-17), P. *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide, Second Edition.* Gerald Farin. Computer Sci. & Scientific Computing. Academic Pr, 1990, xvii + 444 pp, $39.95. [ISBN: 0-12-249051-7] This edition adds programs in C to implement basic methods. (*First Edition,* TR, June-July 1989.) OJ

Computer Graphics, P. *Computation of Curves and Surfaces.* Eds: Wolfgang Dahmen, Mariano Gasca, Charles A. Micchelli. NATO ASI Ser. C, V. 307. Kluwer Academic, 1990, ix + 536 pp, $149. [ISBN: 0-7923-0724-0] Proceedings of a NATO Advanced Study Institute at Puerto de la Cruz, Tenerife, Spain in July 1989. Fifteen papers in the areas of discrete methods for curve and surface representation, interpolation and data fitting, multivariate splines and applications, and algebraic and differential geometric techniques for machine representation of curves and surfaces. OJ

Theory of Computation, P. *Lecture Notes in Computer Science-429: Design and Implementation of Symbolic Computation Systems.* Ed: A. Miola. Springer-Verlag, 1990, xii + 284 pp, $24.70 (P). [ISBN: 0-387-52531-9] About 40 brief papers on aspects of design and implementation of symbolic computation—theory, languages, software environments, and architectures—comprise the proceedings of the DISCO '90 symposium held in April 1990 at Capri, Italy. PZ

Artificial Intelligence, P. *A Computational Model of Metaphor Interpretation.* James H. Martin. Perspectives in Artif. Intellig., V. 8. Academic Pr, 1990, xxiii + 229 pp, $34.95. [ISBN: 0-12-474730-2] Presents a theory of metaphor acquisition, interpretation, and denotation, and describes in detail a computer program applying this theory to natural language interpretation systems. JO

Artificial Intelligence, S(16-18), P, L. *The Handbook of Artificial Intelligence, Volume IV.* Eds: Avron Barr, Paul R. Cohen, Edward A. Feigenbaum. Addison-Wesley, 1989, xv + 699 pp, (P). [ISBN: 0-201-51731-0] Eight chapters, each by different authors. Each chapter is on an independent topic and stresses issues, principles, theory, and the relationship to artificial intelligence. Some topics are repeats from previous volumes but all represent new, fresh views of the field as it is today. Topics: blackboard systems, cooperative distributed problem solving, fundamentals of expert systems, natural language understanding, knowledge-based software engineering, qualitative physics, knowledge-based simulation, computer vision. Bibliography; name and subject indexes. RJA

Artificial Intelligence, P. *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology.* Randall D. Beer. Perspectives in Artif. Intellig., V. 6. Academic Pr, 1990, xxiii + 213 pp, $29.95. [ISBN: 0-12-084730-2] Criticizes the traditional artificial intelligence notion that intelligent computer systems should be modeled after human reasoning, and argues that animal adaptation to environments provides a better model. To back up this argument, the author describes the development of an artificial insect and discusses its behavior. JO

Artificial Intelligence, S. *Common LISP Programming for Artificial Intelligence.* Tony Hasemer, John Domingue. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xiii + 444 pp, (P). [ISBN: 0-201-17579-7] Lisp is a language based on the functional programming model. It is widely used to solve problems in artificial intelligence such as theorem proving, natural language understanding, and expert systems. The dialect called *Common Lisp* is the most widely accepted and used version of the language. This text teaches that language and shows how it can be used to solve problems in the area of artificial intelligence. GMS

Artificial Intelligence, P. *Machine Vision for Three-Dimensional Scenes.* Ed: Herbert Freeman. Academic Pr, 1990, xi + 419 pp, $59.95. [ISBN: 0-12-266722-0] Proceedings of a April 1989 conference sponsored by the Center for Computer Aids for Industrial Productivity (CAIP) of Rutgers University, New Brunswick, New Jersey. Fourteen chapters include twelve papers on various approaches to machine recognition of 3-D objects, and two summaries of panel discussions on the present state of the field. OJ

Artificial Intelligence, S(18), P, L. *Neural Network Models in Artificial Intelligence.* Matthew Zeidenberg. Ser. in Artif. Intellig. Ellis Horwood (US Distr: Wiley), 1990, 268 pp. [ISBN: 0-13-612185-3] Overview of the current state of neural networks research. Exposition is clear enough to make this book a good starting point for learning about the field. Chapters on applications to learning, expert systems, knowledge representation, speech synthesis, visual perception, and language understanding. Good sized bibliography. JO

Artificial Intelligence, T, S. *The Computing Neuron.* Eds: Richard Durbin, Christopher Miall, Graeme Mitchison. Computat. & Neural Systems Ser. Addison-Wesley, 1989, xiv + 417 pp, $44.25. [ISBN: 0-201-18348-X] A collection of twenty-one research papers presented at a meeting entitled "The Neuron as a Computational Unit" held at Cambridge University in 1988. The papers describe recent research work in artificial intelligence, cognitive psy-

chology, and connectionist modelling. GMS

Artificial Intelligence, P. *The Organization of Learning.* C.R. Gallistel. MIT Pr, 1990, xi + 648 pp, $45. [ISBN: 0-262-07113-4] Long discussion of what sorts of spacial and temporal data animals can represent in their nervous systems, and what sorts of computational methods they use to interpret the data. This book is dense with examples of animal behavior and makes fascinating reading. The mathematics in the book is spread out and sometimes hard to find. JO

Computer Science, P. *Formal Semantics and Pragmatics for Natural Language Querying.* James Clifford. Tracts in Theoretical Comput. Sci., V. 8. Cambridge University Pr, 1990, xv + 194 pp, $29.95. [ISBN: 0-521-35433-1] The author argues that existing database theories ignore the passage of time and the corresponding changes in data, so that databases "forget" what they once knew, and are unable to answer simple time-related questions. This book develops the notion of a historical database to address these limitations. JO

Computer Science, T(16-17: 2), S. *Software Engineering: Methods and Management.* Anneliese von Mayrhauser. Academic Pr, 1990, xxviii + 864 pp, $49.95. [ISBN: 0-12-727320-4] Part I: methods and techniques for developing software (problem definition, requirements, specifications, design, coding, testing, operation, and maintenance). Part II: introduction to the management of software engineering projects (management by metrics, feasibility, planning, personnel, leadership, development guidelines). LCL

Applications, P, L. *Fractals in the Natural Sciences.* Eds: M. Fleischmann, D.J. Tildesley, R.C. Ball. Princeton University Pr, 1990, 200 pp, $19.50 (P); $50. [ISBN: 0-691-02438-3; 0-691-08561-7] A collection of papers, each followed by brief questions and answers from a 1988 Royal Society "discussion meeting." Opening paper is by B. Mandelbrot on fractal geometry; others are on applications to physical or life sciences. First published in 1989 by the Royal Society of London. LAS

Applications (Biological Science), P, L. *Mathematical Evolutionary Theory.* Ed: Marcus W. Feldman. Princeton University Pr, 1989, x + 341 pp, $19.95 (P). [ISBN: 0-691-08503-X] A collection of fourteen papers written by students or colleagues of Samuel Karlin to commemorate Karlin's numerous contributions to mathematical evolution theory. Genetics and randomness play central roles in all these papers. The book is divided into two sections. The first section contains papers concerned with general genetic theory, while the second section contains applications of the general theory to specific populations. SM

Applications (Biological Science), S(16-17), P, L. *Fundamentals of Mathematical Evolutionary Genetics.* Yuri M. Svirezhev, Vladimir P. Passekov. Math. & Its Applic., V. 22. Kluwer Academic, 1990, xvi + 395 pp, $129. [ISBN: 90-277-2772-4] Deterministic and stochastic models of selection, mutation, migration, and other factors in evolutionary genetics. Note the price! BC

Applications (Biological Science), P, L. *Lecture Notes in Biomathematics-81: Mathematical Approaches to Problems in Resource Management and Epidemiology.* Eds: C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker. Springer-Verlag, 1989, vii + 327 pp, $33 (P). [ISBN: 0-387-51820-7] Proceedings from a conference held at the Center of Applied Mathematics at Cornell University in Fall 1987. Excellent survey of the current status of applying quantitative methods, some stochastic, to problems concerning population growth, from individual cells to large vertebrates. MK

Applications (Biological Science), S(16-17), P, L. *Population Harvesting: Demographic Models of Fish, Forest, and Animal Resources.* Wayne M. Getz, Robert G. Haight. Mono. in Population Biology, V. 27. Princeton University Pr, 1989, xv + 391 pp, $15.95 (P); $45. [ISBN: 0-691-08516-1; 0-691-08515-3] Fish, animal, and forest resources share the quality of reproducing seasonally, creating large numbers of individuals of a specific age. This age class structure of important populations leads to important similarities in mathematical models describing these populations. This work explores these similarities, and also introduces the reader to nonlinear models. SM

Applications (Economics), S, P, L*. *The Invisible Hand: Economic Equilibrium in the History of Science.* Bruna Ingrao, Giorgio Israel. Transl: Ian McGilvray. MIT Pr, 1990, xiii + 491 pp, $47.50. [ISBN: 0-262-09028-7] An exploration of the history of efforts to mathematicize economics, rooted in the eighteenth century efforts to imitate Newtonian mechanics, centered on the nineteenth century work of Léon Walras which established the idea of equilibrium as the central issue in economic theory, and moving through the twentieth century work of Arrow, Debreau, and Smale. Focuses throughout on the "invariant paradigmatic core" of *existence* (is there a state of compatibility in a competitive market?), *uniqueness* (is there one such state?), and *stability* (do market forces lead to this state?). The answers finally emerged in the last two decades: usually, seldom, and hardly ever. LAS

Applications (Economics), S(17-18), P, L*. *Toward a Formal Science of Economics: The Axiomatic Method in Economics and Econometrics.* Brent P. Stigum. MIT Pr, 1990, xiv + 1033 pp, $45. [ISBN: 0-262-19284-5] Encyclopedia dealing with foundations for a formal methodological basis of econometric theory. Begins with axiomatic methods and mathematical logic. Proceeds to econometric theory of consumer choice, probability theory, nonstandard analysis, and epistemological concerns. Concludes with applications to empirical analysis of econometric theories; determinism, uncertainty, and the utility hypothesis; prediction, distributed lags, and stochastic difference equations. MK

Applications (Engineering), T(14: 1). *Solving Problems in Structures, Volume 2.* P.C.L. Croxton, L.H. Martin. Longman Scientific & Technical (US Distr: Wiley), 1990, vii + 282 pp, $34.95 (P). [ISBN: 0-582-02355-6] Volume 2 of a two-volume set which is part of a series on the solution of problems for engineering students in universities, polytechnics, and colleges. Informal treatment (no proofs) of relevant theory. Worked out and outlines of solutions of problems dealing with the elastic analysis of a hyperstatic structure. Volume 1 dealt with statics and statically determinate structures. JDEK

Applications (Fluid Dynamics), P. *Fluid Dynamics of Viscoelastic Liquids.* Daniel D. Joseph. Appl. Math. Sci., V. 84. Springer-Verlag, 1990, xvii + 755 pp, $59.95. [ISBN: 0-387-97155-6] A study of the elasticity of liquids and of asymptotic theories of constitutive models. Intended as a reference for special topics which do not appear in any other reference. JO

Applications (Fluid Dynamics), P. *Two Phase Flows and Waves.* Eds: Daniel D. Joseph, David G. Schaeffer. IMA, V. 26. Springer-Verlag, 1990, xi + 164 pp, $25. [ISBN: 0-387-97293-5] Eleven papers based on a symposium held at the Institute for Mathematics and Its Applications at the University of Minnesota in January 1989. Focus is on the properties of materials which consist of many small solid particles or grains, and on the development of waves in flowing composites. JDEK

Applications (Physical Science), S(17-18), P. *Computational Mechanics of Nonlinear Response of Shells.* Eds: W.B. Krätzig, E. Oñate. Ser. in Computational Mechanics. Springer-Verlag, 1990, viii + 405 pp, $93. [ISBN: 0-387-52035-X] Unlike most collections of scientific papers, this one represents a careful selection, providing a rigorous introduction based on continuum mechanics and then several advanced papers concerned with new finite element derivations for nonlinear shell problems. MU

Applications (Physical Science), P. *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models.* P. Érdi, J. Tóth. Princeton University Pr, 1989, xxiv + 259 pp, $69.50. [ISBN: 0-691-08532-3] Written jointly by a chemist and a mathematician, this volume surveys mathematical models, both deterministic and stochastic, of chemical kinetics. The last chapter is on applications. Should appeal to chemists with a mathematical inclination. KS

Applications (Physics), T(14: 2), S, L. *A Course in Mathematics for Students of Physics: 2.* Paul Bamberg, Shlomo Sternberg. Cambridge University Pr, 1990, xvii + 443 pp, $54.50. [ISBN: 0-521-33245-1] In this second of two volumes, the authors present some relatively advanced mathematical topics in natural physical contexts. For example, the beginnings of algebraic topology and the calculus of exterior differential forms appear in discussions of electrical networks and electrostatics respectively. Very interesting insights are available to the physically-minded mathematician, but this course is not for the average or the lazy undergraduate. JO

Applications (Physics), P. *Convexity Methods in Hamiltonian Mechanics.* Ivar Ekeland. Ser. of Modern Surveys in Math., Band 19. Springer-Verlag, 1990, x + 247 pp, $79.50. [ISBN: 0-387-50613-6] A thorough treatment of nonlinear systems with Hamiltonian $H(t, p, q)$ convex in the p, q variables. The major focus is on nonlinear analysis, critical point theory, and properties of the index of periodic solutions. The author admits that "many, if not most, important Hamiltonians are nonconvex" (and many that are can be treated with other methods), but still feels that the "unity and simplicity" of this approach "more than offsets the loss of scope." You decide. MPR

Applications (Physics), S, L*. *The Mathematics of Projectiles in Sport.* Neville de Mestre. Australian Math. Soc. Lect. Ser., V. 6. Cambridge University Pr, 1990, xi + 175 pp, $22.95 (P). [ISBN: 0-521-39857-6] Everything you ever wanted to know about projectiles—angles, distances, time of flight, envelopes, drag, spin, Magnus effect, wind—derived from military references but applied here to sports projectiles (golf, cricket, basketball, javelin, etc.). Uses vector calculus and dimensionless equations. Numerous exercises; extensive references. LAS

Applications (Physics), S(18), P. *Symplectic Techniques in Physics.* Victor Guillemin, Shlomo Sternberg. Cambridge University Pr, 1990, xi + 468 pp, $29.95 (P). [ISBN: 0-521-38990-9] Paperback, corrected edition of 1984 text (TR, January 1985). Introduces symplectic geometry in the historical context of the study of optics. This introduction is followed by a chapter laying out the basic theoretical results, and three independent chapters on applications to Yang-Mills equations, completely integrable systems, and symplectic homogeneous spaces. JO

Applications (Physics), S(18), P. *Stochastics, Algebra and Analysis in Classical and Quantum Dynamics.* Eds: S. Albeverio, Ph. Blanchard, D. Testard. Math. & Its Applic., V. 59. Kluwer Academic, 1990, xv + 247 pp, $88. [ISBN: 0-7923-0637-6] This volume collected the Proceedings of the Fourth French-German Encounter on Mathematics and Physics which was held in Marseille, France from February 28 to March 4, 1988. MU

Reviewers

RJA: Richard J. Allen, St. Olaf; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; SG: Steven Galovich, Carleton; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

In appreciation ...

At this time of the year we would like to thank those who have refereed papers for the Monthly during the past year. Without their diligent efforts we could not have functioned. – The editors.

Harvey Abbott, Asuman Aksoy, Stanley Alama, Felix Albrecht, Stephanie Alexander, J. Ralph Alexander, William Allard, Steve Althoen, George E. Andrews, Larry Andrews, Thomas Angell, Tom Apostol, Kenneth Appel, Tom Armstrong, Mark Ashbaugh, Richard Askey, Paul Atela, Kendall Atkinson, Tom Banchoff, Steven Bank, Robert Barnhill, Michael Barr, Michael Beeson, Katalin Bencsath, Edward Bender, John Benedetto, Arthur Benjamin, Jon Bentley, Sterling Berberian, David Berg, Vitaly Bergelson, Bruce Berndt, Ted Bick, Patrick Billingsley, David Bindschadler, Richard Bishop, Tibor Bisztriczky, R. Boas, Joseph Bonin, Robert Borden, Nigel Boston, Michael Botsko, William Boyce, John Brillhart, Duane Broline, Ezra Brown, Herbert Brown, Andrew Bruckner, Robert Burckel, Frank Burk, Donald Burkholder, Lynne Butler, Eugenio Calabi, C. David Callan, Rodney Canfield, G. Cargo, Bille Carlson, George Carrier, Robert Carroll, Jesus Castillo, Frank Cater, Thomas Cecil, Gulbank Chakerian, Roger Chalkley, Jal Choksi, Joseph Cima, Courtney Coleman, Matthew Cook, Don Coppersmith, Paul Corazza, Robert Craggs, George Cross, Robert Currier

Everett Dade, John D'Angelo, Martin Davis, Phillip Davis, John Dawson, Jane Day, Mahlon Day, Carl DeBoor, N. G. DeBruijn, William Derrick, John Dettman, Luc Devroye, M. Dickson, Peter Dixon, J. Doob, Larry Dornhoff, David Doster, Underwood Dudley, Hugh Edgar, H. Edwards, Paul Eggermont, Richard Elderkin, Jim Epperson, Paul Erdős, Richard Escobales, Ronald Evans, Charles Eynden, Kenneth Falconer, Michael Filaseta, Nathan Fine, Stephen Fisher, Harley Flanders, Gerald B. Folland, David Foulis, Michael Frame, George Francis, Eduardo Friedman, Wolfgang Fuchs, Stephen Gagola, Jr., David Gale, Fred Galvin, Adriano Garsia, Ira Gessel, Richard Gibbs, Leonard Gillman, Clark Givens, Alan Gluchoff, Irving Gluck, Richard Goldberg, Herman Gollwitzer, Adolph Goodman, Sandy Grabiner, Judith Grabiner, Ronald Graham, James Graham-Eagle, Alfred Gray, Daniel Grayson, Leon Greenberg, Curtis Greene, Andreas Griewank, Jerrold Griggs, Branko Grunbaum, H. N. Gupta

L. R. Haff, Heine Halberstam, Mary-Elizabeth Hamstrom, David Harbater, Hiroshi Haruki, Maurice Heins, Lester Helms, Michael Henle, Douglas Hensley, Ralph Henstock, David Herron, Edwin Hewitt, Nigel Higson, Adolph Hildebrand, Aimo Hinkkanen, Kevin Hockett, Alan Hoffman, Robert Hogg, Roger Horn, Paul Humke, I. Martin Isaacs, J. Isbell, Arieh Iserles, Howard Jacobowitz, Robert James, Gerald Janusz, Richard Jerrard, Carl Jockusch, Bruce Johnson, Charles Johnson, David Johnson, David Joyce, Dan Kalman, Wilfred Kaplan, Irving Katz, Louis Kauffman, Donald Keedwell, Linda Keen, Arnfried Kemnitz, Clark Kimberling, Roger Kirchner, Victor Klee, Daniel Kleitman, Ronald Knill, Fred Kochman, Paul Koosis, Ralph Kopperman, David Kurtz

Jeffery Lagarias, Tsit-Yuen Lam, Ray Langebartel, Peter Lax, Derrick Lehmer, Emma Lehmer, Joseph Lehner, Calvin Long, William Lucas, Frederick Luttman, Jr., Mark Lynch, Russell Lyons, Thomas MacGregor, Hosam Mahmoud, John Mallet, Walter Mallory Maynard Mansfield, Marvin Marcus, David Marker, Mario Martelli, George Martin, Jerold Mathews, R. Daniel Mauldin, Leon McCulloch, Richard McGehee, Catherine McGeoch, Fred McMorris, Christian Meyer, Kenneth Meyer, Paul Meyers, Ernst Michael, Joseph Miles, Anthony Miller, David Minda, Peter Monk, David S. Moore, John Morgan, John Morrison, B. C. Mortimer, David Muller, James Munkres, Kent Nagle, Zuhair Nashed, Richard Nau, Morris Newman, W. Nicholson, Albert Nijenhuis, Zbigniew Nitecki, Ivan Niven, Richard O'Neil, Karl Norton, Joseph O'Rourke, Beresford Parlett, G. W. Peck, N. Tenney Peck, Richard Pfiefer, Alden Pixley, Esteban Poffald, Carl Pomerance, Horacio Porta, Stephen Portnoy, Josef Przytycki, Bruce Reznick, Billy Rhoades, Arthur Robinson, Derek Robinson, Dave Roeder, C. A. Rogers, S. Roman, Joseph Rosenblatt, Ira Rosenholtz, Richard Roth, Bruce Rothschild, Joseph Rotman, Walter Rudin

Bill Sands, Donald Sarason, Jonathan Schaer, Ken Schilling, James Schmerl, Frank Schmidt, Wlater Schneider, Glenn Schober, Lowell Schoenfeld, James Schupp, Richard Scott, David Secrest, John Selfridge, Marjorie Senechal, Jeffrey Shallit, Daniel Shanks, Louis Shapiro, Donald Sherbert, Don Shimamoto, Herb Silverman, Evelyn Silvia, Benjamin Sims, Michael Singer, David Singer, Walter Sizer, Charles Small, Don Small, Lance Small, F. Smith, David Smith, Ernst Snapper, J. Laurie Snell, Jerome Spanier, Richard Stanley, John Steinig, Richard Stoneham, Gilbert Strang, Karl Stromburg, David Styer, Ted Suffridge, Sheila Sundaram, Lajos Takacs, Carsten Thomassen, Michael Townsend, William Trench, Dan Ullman, Stephen Ullom, Jeff Vaaler, Rudolf Vydorny, Sam Wagstaff, Jr., Walter Wallis, John Walter, Edward Wang, Larry Washington, David Watkins, Paul Weichsel, Elliott Weinberg, Howard Weiner, Gregory Wene, John Wetzel, John Wierman, Tom Wieting, Robert Wijsman, Donald Wilken, K. S. Williams, Robert Woodrow, Patrick Wotus, Mary Wright, D. B. Zagier, Doron Zeilberger, Paul Zorn

Houghton Mifflin 1991

New
Basic Mathematics with Business Applications
Joanne S. Lockwood, Plymouth State College
Richard N. Aufmann and **Vernon C. Barker**
Both of Palomar College
About 200 perforated pages • paperback • Just published

New
Essential Mathematics with Applications, Third Edition
Vernon C. Barker and **Richard N. Aufmann**
About 300 perforated pages • paperback • Just published

New
Prealgebra: Mathematics for a Variable World
Daniel R. Bach and **Patricia Leitner**
Both of Diablo Valley College
About 550 pages • paperback • Just published

Elementary Algebra with Basic Mathematics
Richard N. Aufmann, Vernon C. Barker, and **Joanne S. Lockwood**
462 pages • paperback • 1989

Beginning Algebra
Robert G. Marcucci, San Francisco State University
Harold L. Schoen, The University of Iowa
478 pages • hardcover • 1990

Intermediate Algebra
Norman L. Siever, Los Angeles Valley College
636 pages • hardcover • 1990

New
Algebra with Trigonometry for College Students
Richard N. Aufmann, Vernon C. Barker, and **Joanne S. Lockwood**
About 700 pages • hardcover • Just published

College Algebra and Trigonometry
768 pages • hardcover • 1990

College Algebra
481 pages • hardcover • 1990

College Trigonometry
303 pages • hardcover • 1990

New
Precalculus
About 700 pages • hardcover • Just published

All by **Richard N. Aufmann, Vernon C. Barker,** and
Richard D. Nation, Jr., Palomar College

Precalculus
Dennis Carrie, Golden West College
686 pages • hardcover • 1990

For adoption consideration, request examination packages from your Houghton Mifflin regional office.

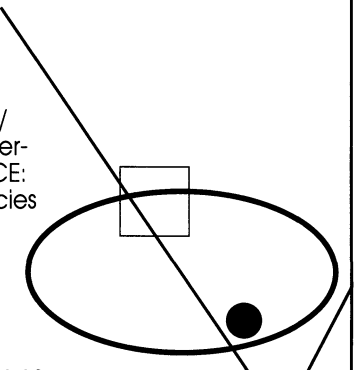
3400 Midway Rd.,
Dallas, TX 75244-5165
1900 S. Batavia Ave.,
Geneva, IL 60134
925 E. Meadow Dr.,
Palo Alto, CA 94303
101 Campus Dr.,
Princeton, NJ 08540

Houghton Mifflin

Building the skills of 3/4 million students . . . and counting

Basic College Mathematics: An Applied Approach, Fourth Edition

About 576 perforated pages • paperback • Instructor's Annotated Edition • Instructor's Resource Manual/ Testing Program • Test Bank • Computerized Test Generator • Computer Tutor™ • Solutions Manual • Math ACE: Additional Computer Exercises • Videos • Transparencies
Just published



Introductory Algebra: An Applied Approach Third Edition

About 576 perforated pages • paperback • Ancillaries as above • Just published

Also available in a hardcover edition.

Intermediate Algebra: An Applied Approach Third Edition

About 656 perforated pages • paperback • Ancillaries as above
Just published

Also available in a hardcover edition.

All by **Richard N. Aufmann** and **Vernon C. Barker**
Both of Palomar College

The program that has helped over three-quarter million students perfect their basic math and algebra skills continues to provide the most comprehensive approach to learning with:

- An interactive approach replete with examples and practice problems *within* the text
- Abundant real-life applications
- An objective-specific format that coordinates text elements and ancillary items.

Now, the best method is even better with:

- New use of full color to highlight important text features
- 100 percent new word problems
- An expanded and improved ancillary package.

For adoption consideration, request examination packages from your Houghton Mifflin regional sales office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
101 Campus Dr., Princeton, NJ 08540

DELLEN AND MACMILLAN


BOOKS!


Developmental Mathematics

Essential Mathematics *1991*
Introductory Algebra 3/e *1991*
Intermediate Algebra 3/e *1991*
all by Linda R. Pulsinelli and Patricia I. Hooper



Basic Mathematics *1991*
Beginning Algebra with Geometry *1991*
Intermediate Algebra *1991*
A Short Course in Geometry *1991*
all by Patricia A. Juelg



Prealgebra *1991*
Developmental Mathematics *1991*
Essential Mathematics *1991*
Basic College Mathematics 3/e *1990*
Elementary Algebra 3/e *1990*
Intermediate Algebra 3/e *1990*
all by Ignacio Bello

Precalculus Mathematics

Trigonometry: An Analytic Approach 6/e *1991*
by Irving Drooyan and Charles C. Carico



Algebra and Trigonometry for College Students *1991*
College Algebra with Review *1991*
College Algebra 2/e *1990*
Algebra and Trigonometry 2/e *1990*
Trigonometry 2/e *1990*
Precalculus 2/e *1990*
all by Michael Sullivan

Applied Mathematics



Applied Mathematics for Business, Economics, Life Sciences, and Social Sciences 4/e *1991*
Applied Calculus for Business, Economics, Life Sciences, and Social Sciences 4/e *1991*
Finite Mathematics for Business, Economics, Life Sciences, and Social Sciences 5/e *1990*
College Mathematics for Business, Economics, Life Sciences, and Social Sciences 5/e *1990*
Calculus for Business, Economics, Life Sciences, and Social Sciences 5/e *1990*
Essentials of College Mathematics: For Business, Economics, Life Sciences, and Social Sciences *1989*
all by Raymond A. Barnett and Michael R. Ziegler

And Keep in Mind...



Supplemental Applications and Topics to Accompany the Barnett/Ziegler College Mathematics Series *1991*
by Jon E. Baum

Add it up with Macmillan and Dellen!

Contact your local representative for more information.

Where Two Plus Two Equals More!

Mathematics for Teachers

Mathematics for Elementary Teachers 2/e *1991*
by Gary L. Musser and William F. Burger



Problem Solving with Elementary Mathematics *1991*
by David Gay

Discrete Mathematics

Discrete Mathematics 2/e *1990*
by Richard Johnsonbaugh

Discrete Mathematics *1990*
by Jerrold W. Grossman

Linear Algebra

Elementary Linear Algebra 5/e *1991*
by Bernard Kolman



Linear Algebra with Applications 3/e *1990*
by Steven J. Leon

Simulation/Operations Research

A Course in Simulation *1991*
by Sheldon M. Ross



Student Supplements

Calculus with the Hewlett-Packard Symbol Manipulating Calculators: HP-28S and HP-48SX *1991*
by Lynn E. Garner



Student's Guide to Using Casio Graphics Calculators *1991*
Student's Guide to Using Hewlett-Packard Graphics Calculators *1991*
Student's Guide to Using Texas Instrument Calculators *1991*
all by Joan Ebert Girard

SOFTWARE!

We offer two different testing systems, which are available for many of our mathematics textbooks: **DellenTest**, a computer-generated random testing system for the IBM or Apple II, and **Macmillan Testing Software**, for the IBM or Macintosh. These systems allow you to create and generate flawlessly structured tests, and best of all, they are both exceedingly easy to use. To find out which system is appropriate for your text, please contact your local representative.

MACMILLAN PUBLISHING COMPANY

College Division•866 Third Avenue•New York, NY 10022
COLLIER MACMILLAN CANADA, INC.•1200 Eglinton Avenue East, Suite 200•Don Mills, Ontario M3C 3N1

SUBSCRIBE TO

UME TRENDS

News and reports on Undergraduate Mathematics Education

Keep up with what's happening in Undergraduate Mathematics Education.

UME TRENDS is conducting a subscription drive for its third volume beginning March 1991.

Whether you are receiving it now or not, you must subscribe in order to keep your issues coming.

We must receive a minimum number of subscriptions in order to keep publishing UME TRENDS.

SUBSCRIBE NOW!

Subscriptions are $12 per year for six issues.

Copy or clip the adjoining form and mail it to arrive by February 8, 1991.

UME TRENDS, Volume 3, 1991. Please check the appropriate category and fill in the remaining information.

___ $12 subscriber within the United States ___ $20 subscriber outside the United States

Check Method of Payment
__ Check or money order __ Visa __ MasterCard

Name _____ Card Number _____

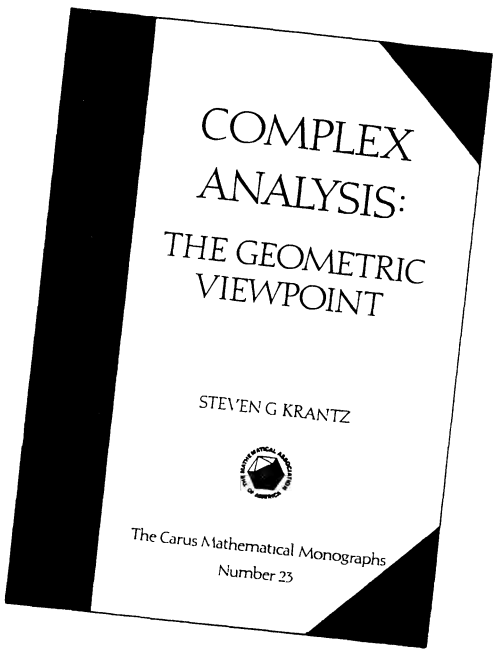
Address _____ Card Expiration date _____

_____ Signature _____

Note that subscriptions must be prepaid.

MAIL TO:
AMERICAN MATHEMATICAL SOCIETY
P. O. Box 1571, Annex Station
Providence, Rhode Island 02901-1571

New From Steven G. Krantz



COMPLEX ANALYSIS:
THE GEOMETRIC VIEWPOINT

Steven G. Krantz

Geometric methods have been used in complex analysis since the 1930s when Lars Ahlfors discovered that they give a nice way to look at the Schwarz lemma. Since that time they have become a central part of the research activities of complex analysis. However these important techniques have never found their way into a text accessible to a broad audience.

Steven G. Krantz, a leading worker in complex analysis and a well-known mathematical expositor, has written the first book explaining how complex analysis can be studied using methods of geometry. Assuming no background in Riemannian geometry, and only one semester of complex analysis, Krantz explains the role of Hermitian metrics and of curvature in understanding the Schwarz lemma, normal families, Picard's theorems, conformal mappings, and many other topics. A minimum of geometric formalism is used to gain a maximum of geometric and analytic insight. The climax of the book is an introduction to several complex variables from the geometric viewpoint. Poincaré's theorem, that the ball and bidisc are biholomorphically inequivalent, is discussed and proved.

Except for the minimal background requirements, the book is self-contained. A review of relevant topics in the classical theory of one complex variable is provided. The style is light and inviting. The book is a must for anyone with an interest in complex analysis. Take a glance at the main chapter headings and order your copy today

- Principal Ideas of Classical Function Theory,
- Basic Notions of Differential Geometry,
- Curvature and Applications,
- Some New Invariant Metrics,
- A Glimpse of Several Complex Variables

210 pp., 1990, Hardbound,
ISBN 0-88385-026-5

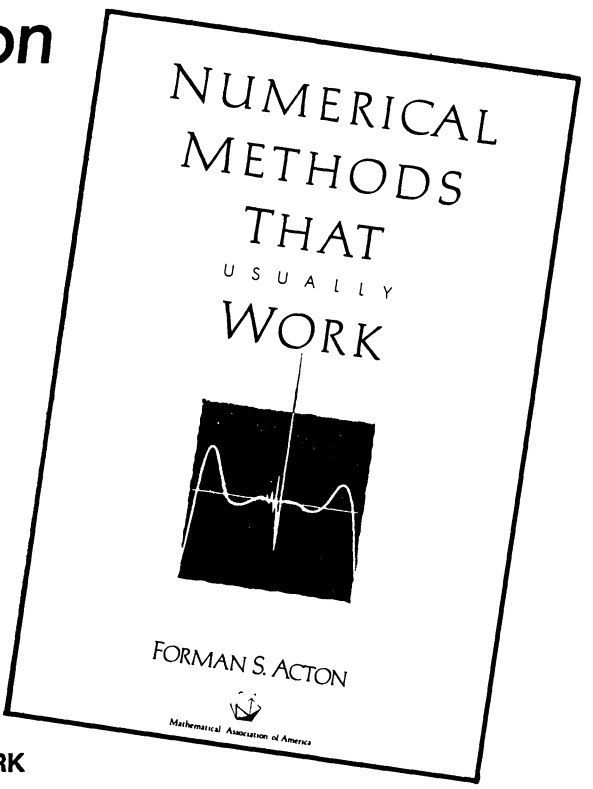
List $22.00 MAA Member $18.50

Catalog Number CAM-23

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

New From Forman S. Acton



NUMERICAL METHODS THAT WORK

Forman Acton

Numerical Methods That Work, originally published in 1970, has been reissued by the MAA with a new preface and some additional problems. Acton deals with a common sense approach to numerical algorithms for the solution of equations: algebraic, transcendental, and differential. It assumes that a computer is available for performing the bulk of the arithmetic. The book is divided into two parts, either of which could form the basis of a one-semester course in numerical methods. Part I discusses most of the standard techniques; roots of transcendental equations, roots of polynomials, eigenvalues of symmetric matrices, and so on. Part II cuts across the basic tools, stressing such common problems as instabilities in extrapolation, removal of singularities, and loss of significant figures. The book is written with clarity and precision, intended for practical rather than theoretical use.

An eminently readable and very well movitivated introductory text.
Mathematics of Computation

A first-rate book which can be used either as a text or reference.
Choice

560 pp., Paperbound, 1990
ISBN 0-88385-450-3

List $16.50 MAA Member $13.00

Catalog Number NMTW

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

University of Arizona
GRADUATE FELLOWSHIPS
IN THE
MATHEMATICAL SCIENCES

Up to 12 fellowships for outstanding new graduate students in the mathematical sciences will be available in 1991-92. Fellowship applicants should be seeking the Ph.D. and planning careers in teaching and/or fundamental research. Anticipated stipends are $12,000 for 12 months with both in-state and non-resident tuition waived.

These fellowships are restricted to citizens, permanent residents, or individuals who have established intent to become citizens or permanent residents of the United States. Applications from U.S. women and students belonging to U.S. minority groups are particularly invited. Currently one-fourth of the U.S. graduate students in pure and applied mathematics at Arizona are women.

The University of Arizona has excellent programs in traditional pure and applied mathematics, and is a leading institution in interdisciplinary applied mathematics. This presents a wealth of opportunities for graduate study encompassing such areas as dynamical systems, number theory, computational science, geometry, nonlinear partial differential equations, mathematical physics, probability, and applications of mathematics in the physical, biological, social, and engineering sciences. In addition, outstanding computational facilities for graduate study and research are available to the over 170 graduate students in the mathematical sciences at the University of Arizona.

Fellowship applicants of superior quality will be among the students invited to the **Fifth Annual Workshop for Advanced Undergraduates on Current Ideas in Nonlinear Science, March 2-5, 1991. Limited support is available for attendees. (The deadline for Workshop applications is February 1, 1991.)** The workshop is designed to communicate topics in current active research in three areas: (i) Geometry in Analysis and Number Theory, (ii) Application of Computers to Mathematical Research, and (iii) Mathematical Applications in the Life Sciences.

For information and application materials contact:
W. M. Greenlee or T. W. Secomb
Department of Mathematics/Program in Applied Mathematics
University of Arizona - Tucson, AZ 85721 - (602) 621-2068
The University of Arizona is an Equal Opportunity Employer

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.50 MAA Member: $5.00

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

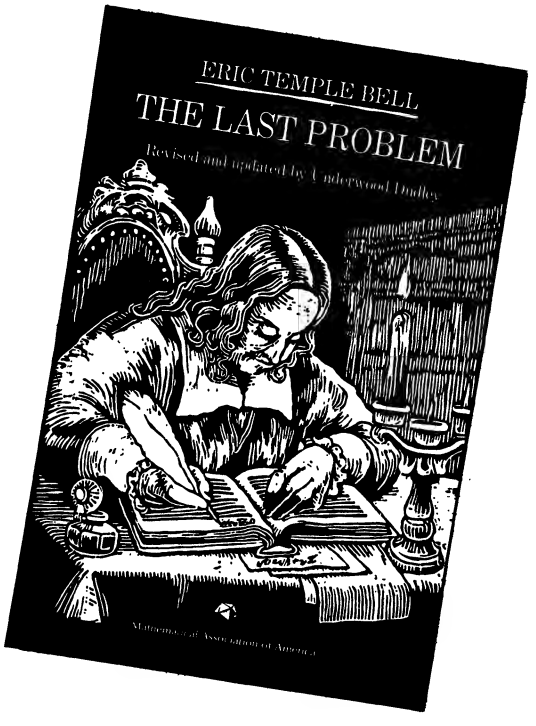
In Writing Mathematics Well, Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, Writing Mathematics Well should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

Revised and Updated



THE LAST PROBLEM

E. T. Bell
Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990
ISBN-0-88385-451-1

List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

HBJ Brings you Range, Depth, and Innovation in 1991

INTRODUCTORY ALGEBRA
and
INTERMEDIATE ALGEBRA
RONALD HATTON and GENE R. SELLERS
both of Sacramento City College

Now in Hardcover!
ELEMENTARY ALGEBRA
Second Edition
and
INTERMEDIATE ALGEBRA
Second Edition
YOSHIKO YAMATO and
MARY JANE CORDON
both of Pasadena City College

FINITE MATHEMATICS
WILLIAM OWEN and FRED CUTLIP
both of Central Washington University

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN, *Drexel University*
and ARNOLD SHAPIRO

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.,
Michigan State University

CALCULATOR ENHANCEMENT FOR A COURSE IN LINEAR ALGEBRA
DON LA TORRE, *Clemson University*

CALCULUS
One and Several Variables
ROBERT ELLIS and DENNY GULICK
both of University of Maryland
Just $29.95 (Net)

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of Vassar College

CALCULATOR ENHANCEMENT FOR A COURSE IN SINGLE VARIABLE CALCULUS
JOHN KENNELLY and
JAMES H. NICHOLSON
both of Clemson University

CALCULATOR ENHANCEMENT FOR A COURSE IN MULTIVARIABLE CALCULUS
JAMES A. RENEKE, *Clemson University*

CALCULATOR ENHANCEMENT FOR A COURSE IN DIFFERENTIAL EQUATIONS
THOMAS G. PROCTOR,
Clemson University

CALCULATOR ENHANCEMENT FOR A COURSE IN INTRODUCTORY STATISTICS
IRIS FETTA, *Clemson University*

ABSTRACT ALGEBRA
DENNIS KLETZING, *Stetson University*



Harcourt Brace Jovanovich, Inc.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

Two Books On Mathematical Ingenuity By Ross Honsberger

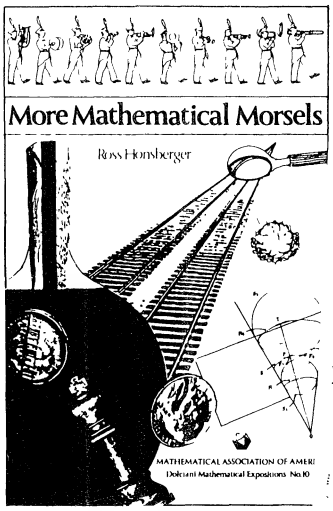
Ross Honsberger is the author of seven books in the Dolciani Mathematical Exposition series, each of which presents problems from algebra, arithmetic, number theory, probability, and geometry, and provides ingenious solutions and/or intriguing results. His most recent addition to the series is MORE MATHEMATICAL MORSELS which is a continuation of the earlier Mathematical Morsels volume published in 1979. Now is your opportunity to own both of these excellent books. The problems presented are meant to be enjoyed, rather than instruct, although instruction is almost always the automatic by-product.

Honsberger says in the Preface of his first volume of morsels, "Mathematics abounds in bright ideas. No matter how long and hard one pursues her, mathematics never seems to run out of exciting surprises. And by no means are these gems to be found only in difficult work at an advanced level. All kinds of simple notions are full of ingenuity. The present volume discusses scores of elementary problems and contains dozens of marvellous ideas."

All of the problems are accessible to anyone with a knowledge of freshman mathematics. Buy both of these excellent books today, and share them with your students.

This is an excellent text...It exposes gifted undergraduate students to a set of elementary problems...whose proofs are ingenious, usually containing beautiful ideas...Its lucid expository style and stimulating mathematical content make its reading a rewarding experience for mathematicians at all levels. (about Mathematical Morsels)

M.S. Cheema in *Mathematical Reviews*



...even the more advanced theor will find many of the problems fa ing...

D.T. Swift-Hook in *Physics Educatic (about Mathematical Morsels)*

NEW NEW NEW NEW NEW N...

MORE MATHEMATICAL MORSELS

Ross Honsberger

315 pp., Paperbound, 1990,

List: $16.00 MAA Member: $13.50

Catalog Number DOL-10

AN MAA BESTSELLER

MATHEMATICAL MORSELS

Ross Honsberger

249 pp., Hardbound, 1979
ISBN 0-88385-303-5

List: $ 28.00 MAA Member: $ 21.00

Catalog Number DOL-03

Both volumes purchased now

List: $40.00 MAA Member: $29.00

Catalog Number DOL-310

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 2 February 1991

Contents

(ISSN 0002-9890)

ARTICLES

Carlyle Circles and the Lemoine Simplicity of Polygon Constructions DUANE W. DeTEMPLE 97

Part Metric and Hyperbolic Metric H. S. BEAR 109

NOTES

Arcs with Positive Measure and a Space-Filling Curve TIMOTHY LANCE AND EDWARD THOMAS 124

Unions and Common Complements of Subspaces K. P. S. BHASKARA RAO AND A. RAMACHANDRA RAO 127

The Generation of All Rational Orthogonal Matrices HANS LIEBECK AND ANTHONY OSBORNE 131

Hypocycloids, Continued Fractions, and Distribution Modulo One NORMAN RICHERT 133

The Perimeter of a Rose JAMES P. BUTLER 139

THE TEACHING OF MATHEMATICS

A Tree Among the Stepping Stones A. R. MEIJER 144

An Analytic Technique to Prove Borel's Strong Law of Large Numbers LIU WEN 146

A Topological Mean Value Theorem for the Plane IRA ROSENHOLTZ 149

Disks and Shells Revisited WALTER CARLIP 154

L'Hôpital's Rule Via Integration DONALD HARTIG 156

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 158

Advanced Problems and Solutions 168

REVIEWS

The Mathematics of Projectiles in Sport. By Neville de Mestre VICTOR G. SZEBEHELY 178

A Transition to Advanced Mathematics, by Douglas Smith, Maurice Eggen, and Richard St. Andre. Foundations for Advanced Mathematics, by Carol Avelsgaard. Introduction to Advanced Mathematics, by William Barnier and Norman Feldman JUDITH PALAGALLO 179

TELEGRAPHIC REVIEWS 182

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $32 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

Carlyle Circles and the Lemoine Simplicity of Polygon Constructions

DUANE W. DETEMPLE, *Washington State University*

DUANE W. DETEMPLE: I received my Ph.D. from Stanford University in 1970, where I wrote my dissertation on bounded univalent functions under the supervision of M. M. Schiffer. I have been at Washington State University since then, with the exception of the 1981–82 year when I was a visiting associate professor at Claremont Graduate School. My current research areas include combinatorial geometry and graph theory.



1. Historical Comments and Introduction. Until very late in the eighteenth century the only regular polygons known to have Euclidean straightedge and compass constructions were precisely the ones shown in Book IV of Euclid's *Elements*. This situation changed abruptly on March 30, 1796, when one month before his nineteenth birthday Carl Friedrich Gauss made the first entry in his notebook [7]:

> *Principia quibus innititur sectio circuli, ac divisibilitas geometrica in septemdecim partes etc.*
>
> *Mart.* 30 *Bruns* [*igae*].
>
> *Principle of the circle's division, and how one geometrically divides the circle into seventeen parts, and so forth.*
>
> *March* 30 [*Braunschweig*]

Gauss had discovered that besides the regular polygons of $2^n \cdot 3$, $2^n \cdot 4$, $2^n \cdot 5$ and $2^n \cdot 15$ sides, there were a number of other constructible polygons, including the 17-gon. Gauss secured priority to his discovery by publishing an announcement on June 1, 1796, which appeared in the "Intelligenzblatt der Allgemeinen Literaturzeitung," the first and only time he published in a journal of advance notices. In his short note he wrote that "This discovery is really only a corollary of a theory with greater content, which is not complete yet, but which will be published as soon as it is complete" ([1], [7]). The full meaning of Gauss' pronouncement came in 1801 with the publication of his monumental *Disquisitiones Arithmeticae* [6]. Its final two sections (articles 365 and 366) discuss the issue of polygon constructibility. Recasting the problem in terms of constructing the N vertices of the polygon on a given circle, Gauss stated his result as follows.

> In general therefore in order to be able to divide the circle geometrically into N parts, N must be 2 or a higher power of 2, *or* a prime number of the form $2^m + 1$, or the product of several prime numbers of this form, *or* the product of one or several such primes into 2 or a higher power of 2.

The sufficiency of the condition follows readily from Gauss' analysis. The necessity, however, is not obvious and Gauss never published a proof of this assertion. The first proof is credited to Pierre L. Wantzel (1814–48) [15].

Primes of the form $2^m + 1$, and therefore necessarily of the form $F_k = 2^{2^k} + 1$, are the Fermat primes. Gauss knew that $F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$ and $F_4 = 65537$ were prime, and he knew that Euler had shown F_5 to be composite. Even today no other Fermat primes have been found, although the smallest unsettled cases are F_{22}, F_{24}, and F_{28}. (The primality status of Fermat numbers as of 1983 can be found in Keller [**8**], and a shorter but more recent table is included in Young and Buell [**17**]).

Beyond the brilliant theoretical breakthrough of Gauss there is still an intriguing puzzle: how can one devise a sequence of steps with straightedge and compass which constructs a regular N-gon? Gauss expressed interest in this problem, although he did not offer explicit geometric constructions. Throughout the last century, and even into present times, numerous constructions have been contrived for the 17-gon (see [**1**], [**2**] for historical comments). J. F. Richelot's construction of the 257-gon required a total of 194 pages, published in four parts in Crelle's journal in 1832. A Professor O. Hermes labored for ten years on the construction and associated algebra of the 65537-gon. The work filled a trunk which was donated to the Mathematical Institute at Göttingen. Nearly a century later it remains stored in an attic there, in all likelihood having never been read.

In what follows we show that a remarkably uniform procedure to construct the regular F_k-gons is afforded by consistent use of an idea attributed to the Scottish historian Thomas Carlyle (1795–1881). Before turning his attention to the literary arts, the young Carlyle taught mathematics, translated Legendre's influential *Elements de Géométrie* into English, and, important for our purposes, devised an elegant geometrical solution to quadratic equations. His idea is based on what we will call the Carlyle circle construction, which we describe in the next section. In the remaining sections we will demonstrate constructions of the 5-, 17- and 257-gon which employ the Carlyle circle method, and, at least in a general way, we also discuss the method's applicability to the 65537-gon.

Our unified approach has two nice features: first, the correctness of the constructions is easy to verify, and second, the constructions are highly efficient. For the pentagon the procedure is similar to the popular construction from antiquity found in Ptolemy's *Almagest*, but with a slight twist at the end. For the 17-gon we rediscover a method of Smith [**12**] which appeared in 1920, but again we can offer a small improvement. The 257-gon is fairly difficult at the concrete level of actually using the Euclidean tools with sufficient accuracy to produce a convincing construction, but at the conceptual level the procedure is remarkably straightforward in its use of 24 Carlyle circles. Indeed, one becomes convinced that the style of construction needs no change to handle the 65537-gon.

A quantitative measure of the simplicity of a geometric construction was devised in 1907 by Émile Lemoine (1840–1912), and is described in Eves [**5**] this way. Consider the following five operations.

S_1: to make the straightedge pass through one given point.

S_2: to rule a straight line.

C_1: to make one compass leg coincide with any point.

C_2: to make one compass leg coincide with any point of a given locus.

C_3: to describe a circle.

If these operations are performed m_1, m_2, n_1, n_2, n_3 times respectively in a construction then $m_1S_1 + m_2S_2 + n_1C_1 + n_2C_2 + n_3C_3$ is the *symbol* of the construc-

tion. The total number of operations $m_1 + m_2 + n_1 + n_2 + n_3$ is called the *simplicity* of the construction, where we note that a low value of simplicity corresponds to an efficient construction. In what follows we will assume that we use a single noncollapsing compass[1]. Simplicity is improved by having circles with a common radius, or a common center, described in succession since the compass would not have to be reset, or one of its legs would not require repositioning.

As an example, the construction of a circle and two lines which cross orthogonally at the circle's center (if we begin with the circle, C_3, then draw a line through the circle's center, $S_1 + S_2$, and finally erect the perpendicular bisector, $2C_1 + 2C_3 + 2S_1 + S_2$) has symbol $3S_1 + 2S_2 + 2C_1 + 0C_2 + 3C_3$ and simplicity 10. If the first two steps in the construction just described are reversed (begin with a line, S_2, and then draw a circle centered on the line, $C_2 + C_3$) then the symbol is $2S_1 + 2S_2 + 2C_1 + C_2 + 3C_3$, but the simplicity is still 10.

The constructions we will describe for regular polygons have lower simplicity measures than their competitors and are perhaps nearly optimal. An interesting, but apparently open, problem is whether or not it can be proved that a given construction has optimal Lemoine simplicity.

2. Carlyle Circles. A variety of methods to construct the roots of a quadratic equation are known, but the one of Carlyle we now describe is especially attractive. It appeared in Leslie's *Elements of Geometry* with the remark "The solution of this important problem now inserted in the text, was suggested to me by Mr. Thomas Carlyle, an ingenious young mathematician, and formerly my pupil" (see [5]).

Beginning with the usual Cartesian x, y-axes and a unit distance, our intention is to construct the roots of the quadratic equation $x^2 - sx + p = 0$, where s and p are given signed lengths. To this end, plot the points $A(0, 1)$ and $B(s, p)$. The circle which has the segment $\overline{AB}$ as a diameter will be called the *Carlyle circle* $C_{s,p}$ of the given quadratic equation. The center of C_{sp} is at $M(s/2, (1 + p)/2)$, which can be constructed as the midpoint of $\overline{AB}$. It will later be found useful to notice that M is also the midpoint of $S(s, 0)$ and $Y(0, 1 + p)$, where the advantage derives from being able to locate S and Y on the x- and y-axes, respectively.

If we suppose $C_{s,p}$ crosses the x-axis at $H_1(x_1, 0)$ and $H_2(x_2, 0)$, with $x_1 \geqslant x_2$, we have the two cases shown in FIG. 1. For p of either sign we observe that $x_1 + x_2 = s$. For $p \leqslant 0$ the intersecting chords theorem for FIGURE 1(a) shows $OH_1 \cdot OH_2 = OA \cdot OC$; that is, $(x_1) \cdot (-x_2) = (1) \cdot (-p)$. For $p > 0$, the intersecting secants theorem for FIG. 1(b) shows $x_1 \cdot x_2 = p$. In all cases, then,

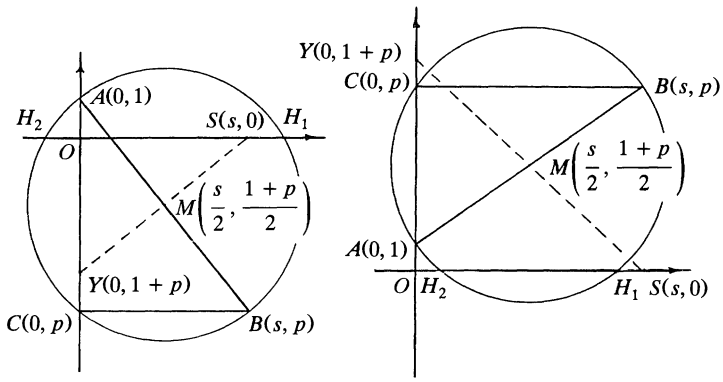
$$(x - x_1)(x - x_2) = x^2 - sx + p, \qquad s = x_1 + x_2, \qquad p = x_1 x_2,$$

and so we conclude the following theorem:

If the Carlyle circle $C_{s,p}$ intersects the x-axis at x_1 and x_2, these are the roots of $x^2 - sx + p = 0$.

An alternate proof can be based on the Pythagorean theorem.

[1]The modern noncollapsing compass permits one to draw with ease a circle centered at a given point whose radius is the distance between two other points. If one only has a collapsing compass, the type assumed by Euclid, then it is surprisingly cumbersome to transfer lengths and consequently the simplicity numbers of most constructions is greatly increased.



(a) Case $p \leqslant 0$ (b) Case $p > 0$

FIG. 1

The quadratic equations we will consider later all have real roots, but it should be pointed out that complex roots are also related to the Carlyle circle. Suppose the y-coordinate $(1 + p)/2$ of the center of $C_{s,p}$ exceeds the circle's radius

$$r = \left[\left(\frac{s}{2}\right)^2 + \left(\frac{1-p}{2}\right)^2\right]^{1/2},$$

a condition which reduces to having a negative discriminant $\Delta = s^2 - 4p$. The circle $C_{s,p}$ does not intersect the x-axis, which reflects the fact that the quadratic equation has a pair of complex conjugate roots $z_1, z_2 = s/2 \pm \frac{1}{2}i\sqrt{-\Delta}$. We will now verify that any circle centered on the x-axis and orthogonal to $C_{s,p}$ will meet the vertical line through the center of $C_{s,p}$ at

$$\left(\frac{s}{2}, \frac{1}{2}\sqrt{-\Delta}\right) \quad \text{and} \quad \left(\frac{s}{2}, -\frac{1}{2}\sqrt{-\Delta}\right).$$

To see this, consider the circle centered at $(\xi, 0)$ which is orthogonal to $C_{s,p}$. Its equation is

$$(x - \xi)^2 + y^2 = \left(\frac{s}{2} - \xi\right)^2 + \left(\frac{1+p}{2}\right)^2 - r^2,$$

and setting $x = s/2$ we find

$$y^2 = \left(\frac{1+p}{2}\right)^2 - r^2 = -\tfrac{1}{4}\Delta.$$

Therefore any such orthogonal circle can be used to construct z_1, and z_2, with the y-axis now interpreted as the imaginary axis of the complex plane.

3. Construction of the Regular Pentagon. We assume that we have already constructed the unit circle and the x, y-axes. Following Gauss we view the problem as the construction of the roots of $z^5 - 1 = 0$. The root $z_0 = 1$, corresponding to the point $P_0(1, 0)$, can be factored out to show that the remaining points P_1, P_2, P_3, P_4 correspond to the roots of the quartic equation $z^4 + z^3 + z^2 +$

$z + 1 = 0$. Letting $\varepsilon = e^{2\pi i/5}$ these roots are $\varepsilon^1, \varepsilon^2, \varepsilon^3, \varepsilon^4$, which we observe sum to -1.

Two other sums of roots are of special interest, namely,

$$\begin{aligned} \eta_0 &= \varepsilon^1 + \varepsilon^4 = 2\cos(2\pi/5) \\ \eta_1 &= \varepsilon^2 + \varepsilon^3 = 2\cos(4\pi/5). \end{aligned} \tag{1}$$

It is simple to check that

$$\eta_0 + \eta_1 = -1, \qquad \eta_0 \cdot \eta_1 = -1, \qquad \eta_0 > 0 > \eta_1. \tag{2}$$

From (2) it follows that η_0 and η_1 are, respectively, the larger and smaller root of the quadratic equation $x^2 + x - 1 = 0$, and therefore they can be constructed by means of the Carlyle circle $C_{-1,-1}$ which has center $M(-1/2, 0)$. M is constructed as the midpoint of $\overline{OQ}$, where $Q(-1, 0)$. The Carlyle circle centered at M and passing through $A(0, 1)$ then intersects the x-axis at $H_0(\eta_0, 0)$ and $H_1(\eta_1, 0)$. In view of (1), circles of unit radius centered at H_0 and H_1 will intersect the original unit circle at P_1, P_2, P_3, P_4, thereby forming the desired pentagon $P_0P_1P_2P_3P_4$ as shown in FIGURE 2.

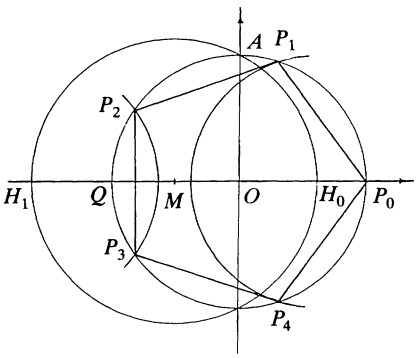


FIG. 2

Beginning with the unit circle and the x, y-axes as given, the construction just described has symbol $2S_1 + S_2 + 8C_1 + 0C_2 + 4C_3$, for a Lemoine simplicity of 15. The popular construction of Ptolemy (see [4], [15]) proceeds in the same way through the construction of H_0, but then takes AH_0 as the distance between successive vertices of the pentagon, resulting in a simplicity of 16 to locate all five vertices.

4. Construction of the Regular Heptadecagon (17-gon). Letting $\varepsilon = e^{2\pi i/17}$, our task here is to construct the roots $\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^{16}$ of the equation $z^{16} + z^{15} + \cdots + z + 1 = 0$. Gauss' idea was to arrange these roots in a cycle in which each root is a gth power of its predecessor:

$$\varepsilon^1, \varepsilon^g, \varepsilon^{g^2}, \ldots, \varepsilon^{g^{15}}, \varepsilon^{g^{16}} = \varepsilon^1. \tag{3}$$

Since $\varepsilon^m = \varepsilon^{m \pmod{17}}$, we can see that corresponding to the prime $p = 17$ there must be found a g for which

$$g^r \not\equiv 1 \pmod{p}, \qquad 1 \leqslant r \leqslant p - 2 \quad \text{and} \quad g^{p-1} \equiv 1 \pmod{p}. \tag{4}$$

A positive integer g which satisfies the conditions of (4) for p is said to be a primitive root of the prime p. For any Fermat prime $p > 3$, $g = 3$ is always a primitive root. This follows from the fact that 3 is a quadratic nonresidue by the quadratic reciprocity theorem, and any nonresidue is a primitive root because the size of the multiplicative group is a power of 2. With this choice the cycle (3) becomes

$$\varepsilon^1, \varepsilon^3, \varepsilon^9, \varepsilon^{10}, \varepsilon^{13}, \varepsilon^5, \varepsilon^{15}, \varepsilon^{11}, \varepsilon^{16}, \varepsilon^{14}, \varepsilon^8, \varepsilon^7, \varepsilon^4, \varepsilon^{12}, \varepsilon^2, \varepsilon^6. \tag{5}$$

Next, Gauss considered the sums of the terms of various subcycles of (5), which are called periods. A typical example is the period $\eta_{0,2} = \varepsilon^1 + \varepsilon^9 + \varepsilon^{13} + \cdots + \varepsilon^2$. For notational convenience a period will be represented by an l-tuple of the exponents of ε which occur in the sum, and so the two periods of length eight are expressed as

$$\begin{aligned} \eta_{0,2} &= (1,9,13,15,16,8,4,2) \\ \eta_{1,2} &= (3,10,5,11,14,7,12,6). \end{aligned} \tag{6}$$

There are four periods of length four, namely,

$$\begin{aligned} \eta_{0,4} &= (1,13,16,4), \qquad \eta_{2,4} = (9,15,8,2) \\ \eta_{1,4} &= (3,5,14,12), \qquad \eta_{3,4} = (10,11,7,6). \end{aligned} \tag{7}$$

There are eight periods of length two, but only two are needed for the construction,

$$\begin{aligned} \eta_{0,8} &= (1,16) = 2\cos(2\pi/17) \\ \eta_{4,8} &= (13,4) = 2\cos(8\pi/17). \end{aligned} \tag{8}$$

From (8), and using (7), we easily see that

$$\eta_{0,8} + \eta_{4,8} = \eta_{0,4}, \qquad \eta_{0,8} \cdot \eta_{4,8} = \eta_{1,4}, \qquad \eta_{0,8} > \eta_{4,8}. \tag{9}$$

Similarly it is straightforward to calculate that

$$\eta_{0,4} + \eta_{2,4} = \eta_{0,2}, \qquad \eta_{0,4} \cdot \eta_{2,4} = -1, \qquad \eta_{0,4} > \eta_{2,4} \tag{10}$$

$$\eta_{1,4} + \eta_{3,4} = \eta_{1,2}, \qquad \eta_{1,4} \cdot \eta_{3,4} = -1, \qquad \eta_{1,4} > \eta_{3,4} \tag{11}$$

$$\eta_{0,2} + \eta_{1,2} = -1, \qquad \eta_{0,2} \cdot \eta_{1,2} = -4, \qquad \eta_{0,2} > \eta_{1,2}. \tag{12}$$

The inequalities in (10) and (11) are geometrically obvious, and since a calculation shows that $(\eta_{0,4} - \eta_{2,4}) \cdot (\eta_{1,4} - \eta_{3,4}) = 2(\eta_{0,2} - \eta_{1,2})$ it then follows that $\eta_{0,2} > \eta_{1,2}$ as claimed in (12).

Equations (8)–(12) readily lead to the construction of the regular 17-gon as shown in FIGURE 3. It is assumed that we are given the unit circle centered at O and the x, y-axes. From (12) we see that $\eta_{0,2}$ and $\eta_{1,2}$ are, respectively, the larger and smaller roots of the quadratic equation $x^2 + x - 4 = 0$ whose corresponding Carlyle circle $C_{-1,-4}$ is centered at $M_0(-1/2, -3/2)$. Therefore:

(i) Draw the perpendicular bisector of $\overline{QO}$, using a circle at Q through O and the existing circle. This construction has Lemoine symbol $2S_1 + S_2 + 2C_1 + C_3$ and locates Q'.
(ii) Draw a circle at Q' through P_0 (symbol $2C_1 + C_3$) to locate M_0.
(iii) Draw the Carlyle circle at M_0 (symbol $2C_1 + C_3$) to locate $H_{0,2}(\eta_{0,2}, 0)$ and $H_{1,2}(\eta_{1,2}, 0)$.

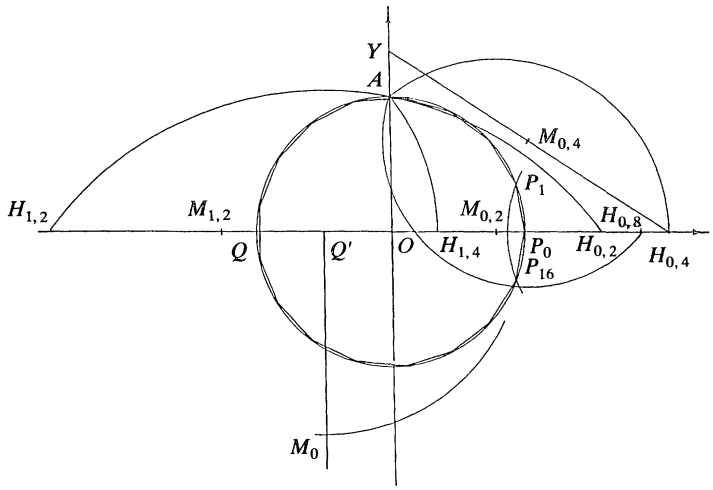


FIG. 3

From equations (10) and (11) we see that Carlyle circles centered at $M_{0,2}(\frac{1}{2}\eta_{0,2}, 0)$ and $M_{1,2}(\frac{1}{2}\eta_{1,2}, 0)$ will locate $H_{0,4}(\eta_{0,4}, 0)$ and $H_{1,4}(\eta_{1,4}, 0)$. Therefore:

(iv) Draw the perpendicular bisectors of $\overline{OH}_{0,2}$ and $\overline{OH}_{1,2}$, using the same circle at O (symbol $4S_1 + 2S_2 + 3C_1 + 3C_3$). This locates $M_{0,2}$ and $M_{1,2}$.

(v) Draw the Carlyle circles at $M_{0,2}$ and $M_{1,2}$ (symbol $4C_1 + 2C_3$) to locate $H_{0,4}$ and $H_{1,4}$.

Turning next to Equation (9), we see that $H_{0,8}(\eta_{0,8}, 0)$ requires a Carlyle circle at $M_{0,4}$. Therefore:

(vi) Set the compass to radius $QH_{1,4}$ and draw a circle at O (symbol $3C_1 + C_3$) to locate $Y(0, 1 + \eta_{1,4})$.

(vii) Draw the line $\overline{YH}_{0,4}$ (symbol $2S_1 + S_2$).

(viii) Draw the perpendicular bisector to $\overline{YH}_{0,4}$ (symbol $2S_1 + S_2 + 2C_1 + 2C_3$) to locate $M_{0,4}$.

(ix) Draw the Carlyle circle at $M_{0,4}$ (symbol $2C_1 + C_3$) to locate $H_{0,8}$.

In view of equation (8), the following step locates P_1 and P_{16}, which correspond to ε^1 and ε^{16}, respectively. Except for laying off the distance P_0P_1, this completes the construction.

(x) Set the compass to unit radius and draw a circle at $H_{0,8}$ (symbol $3C_1 + C_3$) to locate P_1 and P_{16}.

The ten steps described above have total symbol $10S_1 + 5S_2 + 23C_1 + 0C_2 + 13C_3$, for a simplicity of 51. However, two simple modifications can reduce the simplicity of the construction to 45.* First, the costly double bisection of step (iv) can be avoided by using half-scaled Carlyle circles to locate $M_{0,2}$ and $M_{1,2}$. This requires a perpendicular bisector of $\overline{Q'O}$, but since the compass is already set at

*Our thanks to the referee for suggesting these modifications.

unit radius the symbol is $2S_1 + S_2 + C_1 + C_3$. We also need a circle of radius OQ' at O (symbol $2C_1 + C_3$). Since there is no more need for step (iv) the simplicity is reduced by four. Step (viii) can also be replaced by a less costly procedure. Following step (vi) the perpendicular bisector of OY can be constructed with symbol $2S_1 + S_2 + C_1 + C_3$, since the compass need not be reset for the circle at Y. Step (vii) now locates $M_{0,4}$ and so we have further reduced the simplicity by two. Altogether the modified Carlyle construction has symbol $8S_1 + 4S_2 + 22C_1 + 11C_3$, for a simplicity of 45.

The Smith construction [**12**] is similar to that described in (i)–(x), but it requires a perpendicular bisector at $H_{0,4}$, which increases the simplicity to 58. The construction of Richmond [**11**] (also shown in [**13**], [**16**], and elsewhere) seems to appear more often than its efficiency warrants. Even arranging for several circles to do double duty, as in the modified Carlyle construction, the simplicity is 53 (which constructs P_3; the simplicity increases to 60 to locate P_1.) The comparatively recent construction of Tietze [**14**], also described in Hall [**7**], can be done with simplicity 50. Neither reference provides a proof of the construction, and indeed Hall remarks that "the proof that it is correct requires extensive calculation." His assessment applies equally well to the method of Richmond.

5. Construction of the Regular 257-gon. Since $g = 3$ is also a primitive root of 257, the roots of $z^{256} + z^{255} + \cdots + 1 = 0$ are ordered in the cycle

$$\varepsilon^1, \varepsilon^3, \varepsilon^9, \varepsilon^{27}, \varepsilon^{81}, \varepsilon^{243}, \varepsilon^{215}, \ldots, \varepsilon^{165}, \varepsilon^{238}, \varepsilon^{200}, \varepsilon^{86},$$

where $\varepsilon = e^{2\pi i/257}$. The period of length $l = (p-1)/2^r$ which contains the term ε^{g^j} is denoted by

$$\eta_{j,k} = \sum_{n=0}^{l-1} \varepsilon^{g^{j+nk}}, \qquad k = 2^r.$$

In particular, $\eta_{0,1} = -1$ and we also observe that $\eta_{j,k} = \eta_{i,k}$ when $j \equiv i \pmod{k}$.

To set up the hierarchy of quadratic equations to be solved in succession, it is necessary to know the sum, product, and relative size of pairs of periods. This allows the construction of the periods by means of the appropriate Carlyle circles.

The addition formulas

$$\eta_{j,2k} + \eta_{j+k,2k} = \eta_{j,k} \tag{13}$$

are obvious. Formulas for products require effort but perhaps less than expected. The following sample calculation of $\eta_{0,64} \cdot \eta_{32,64}$ illustrates some useful principles of computation. Since $\eta_{0,64} = (1, 241, 256, 16)$ and $\eta_{32,64} = (64, 4, 193, 253)$ are represented by quadruples of exponents, the product we seek corresponds to a modulo 257 addition table.

		$\eta_{32,64}$			
		64	4	193	253
$\eta_{0,64}$	1	65	5	194	254
	241	48	245	177	237
	256	63	3	192	252
	16	80	20	209	12

The main diagonal, it can be checked, is $\eta_{33,64} = (65, 245, 192, 12)$, and the "diagonal" just below it is $\eta_{55,64} = (48, 3, 209, 254)$. Altogether we find that

$$\eta_{0,64} \cdot \eta_{32,64} = \eta_{33,64} + \eta_{55,64} + \eta_{23,64} + \eta_{1,64}. \tag{14}$$

Thus products can be computed by knowing which periods of the same length contain the terms in the first column.

The right side of (14) can be simplified by use of the addition formulas (13), giving

$$\eta_{0,64} \cdot \eta_{32,64} = \eta_{1,32} + \eta_{23,32}. \tag{15}$$

The derivation of (15) is actually enough from which to deduce the general formula obtained by increasing the first subscripts by j. That is,

$$\eta_{j,64} \cdot \eta_{j+32,64} = \eta_{j+1,32} + \eta_{j+23,32}. \tag{16}$$

Additional discussion of these computational details can be found in Rademacher [**10**].

The last algebraic difficulty is the ordering of $\eta_{j,2k}$ and $\eta_{j+k,2k}$. It is obvious that $\eta_{0,64} > \eta_{32,64}$ but showing that $\eta_{4,16} > \eta_{12,16}$ is somewhat delicate. The methods described in Klein [9] and Rademacher [**10**] for the 17-gon can be extended to the present case however, thus allowing the identification of the larger and smaller root of a pair.

The results of the calculations are shown in TABLE 1. The same information is contained in Bishop [**3**], but in a different form, since we have used formula (13) to simplify the expressions for the period products. Since it will be sufficient to construct the point P_1 on the unit circle whose x-coordinate is $\cos(2\pi/257) = (1/2)\eta_{0,128}$, we observe that we need only construct two of the period pairs of length 4 and six of length 8.

TABLE 1

Products of Required Period Pairs	Order of Period Pairs
$\eta_{0,2} \cdot \eta_{1,2} = -64$	$\eta_{0,2} > \eta_{1,2}$
$\eta_{j,4} \cdot \eta_{j+2,4} = -16, \quad j = 0, 1$	$\eta_{0,4} > \eta_{2,4}, \eta_{1,4} > \eta_{3,4}$
$\eta_{j,8} \cdot \eta_{j+4,8} = -2 + 3\eta_{j+1,2} + 2\eta_{j+2,4}$ $j = 0, 1, 2, 3$	$\eta_{0,8} > \eta_{4,8}, \eta_{2,8} > \eta_{6,8},$ $\eta_{1,8} < \eta_{5,8}, \eta_{3,8} < \eta_{7,8}$
$\eta_{j,16} \cdot \eta_{j+8,16} = \eta_{j,2} + \eta_{j,8} + \eta_{j+2,8} + 2\eta_{j+5,8}$ $j = 0, \ldots, 7$	$\eta_{0,16} > \eta_{8,16}, \eta_{4,16} > \eta_{12,16}, \eta_{2,16} > \eta_{10,16},$ $\eta_{6,16} < \eta_{14,16}, \eta_{1,16} > \eta_{9,16}, \eta_{5,16} > \eta_{13,16},$ $\eta_{3,16} > \eta_{11,16}, \eta_{7,16} > \eta_{15,16}$
$\eta_{j,32} \cdot \eta_{j+16,32} = \eta_{j,16} + \eta_{j+1,16}$ $+\eta_{j+2,16} + \eta_{j+5,16}$ $j = 0, 1, 7, 8, 9, 15$	$\eta_{0,32} > \eta_{16,32}, \eta_{8,32} < \eta_{24,32},$ $\eta_{1,32} > \eta_{17,32}, \eta_{9,32} < \eta_{25,32},$ $\eta_{7,32} < \eta_{23,32}, \eta_{15,32} > \eta_{31,32}$
$\eta_{j,64} \cdot \eta_{j+32,64} = \eta_{j+1,32} + \eta_{j+23,32}$ $j = 0, 24$	$\eta_{0,64} > \eta_{32,64}, \eta_{24,64} < \eta_{56,64}$
$\eta_{0,128} \cdot \eta_{64,128} = \eta_{56,64}$	$\eta_{0,128} > \eta_{64,128}$

Translating the results of the table into an explicit construction is straight-forward, although we must be alert to carrying out the required procedure in an efficient manner. Beginning with the first row of information in TABLE 1 we see

that $\eta_{0,2}$ and $\eta_{1,2}$ are the larger and smaller roots, respectively, of $x^2 + x - 64 = 0$, and so the first step is to construct the center $M_0(-1/2, -63/2)$ of the Carlyle circle $C_{-1,-64}$. Successive distance doublings along the positive x-axis will locate $T(31, 0)$, and then swinging an arc from $Q'(-1/2, 0)$ of radius $Q'T$, will locate M_0 on the perpendicular bisector $x = -1/2$ of O and $Q(-1, 0)$. As shown in FIGURE 4, the Carlyle circle $C_{-1,-64}$ then locates $H_{0,2}(\eta_{0,2}, 0)$ and $H_{1,2}(\eta_{1,2}, 0)$.

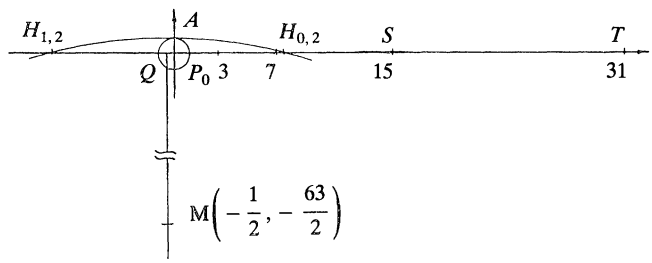


FIG. 4

The construction of the four periods of length 64 is depicted in FIGURE 5. The point $Y(0, -15)$ is located by swinging an arc from O of radius OS, where $S(15, 0)$ was constructed in the previous step. The midpoint $M_{0,2}$ of $\overline{YH}_{0,2}$ is the center of the Carlyle circle which determines $H_{0,4}(\eta_{0,4}, 0)$ and $H_{2,4}(\eta_{2,4}, 0)$. Similarly the points at $x = \eta_{1,4}$ and $x = \eta_{3,4}$ are found by the Carlyle circle centered at the midpoint $M_{1,2}$ of $\overline{YH}_{1,2}$.

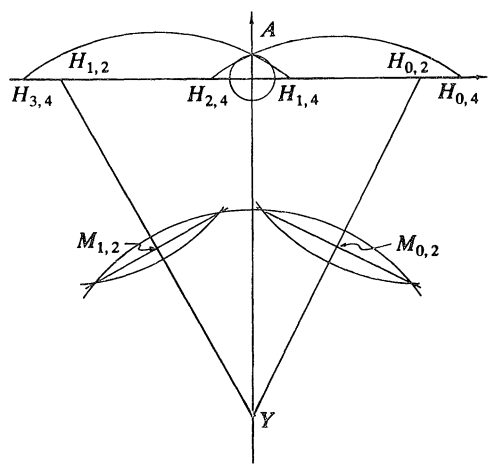


FIG. 5

The remaining pairs of periods are constructed in an analogous way, although extra distance transfers are required because of the more complicated product expressions shown in TABLE 1. A circle of unit radius centered at $H_{0,128}$ will cross the unit circle at P_1 and P_{256}, which completes the construction up to laying off the distance P_0P_1 to locate all of the vertices of the 257-gon.

To optimize the simplicity of the construction it is essential to transfer distances efficiently. For example some useful savings in constructing $\eta_{0,16}, \eta_{4,16}, \eta_{8,16}, \eta_{12,16}$ are possible when advantage is taken of the relation $\eta_{0,16} \cdot \eta_{8,16} + \eta_{4,16} \cdot \eta_{12,16} = 3\eta_{0,2} + 2\eta_{1,4}$. The construction of the three other quadruples of periods of length sixteen are aided by the corresponding algebraic relations.

Beginning with the unit circle and the x, y-axes, the construction of $P_1(\cos(2\pi/257), \sin(2\pi/257))$ has symbol

$$94S_1 + 47S_2 + 275C_1 + 0C_2 + 150C_3$$

for a Lemoine simplicity of 566. Among the 150 circles are 24 Carlyle circles.

6. Remarks on the Construction of the Regular 65537-gon. We have already observed that $g = 3$ is a primitive root of $p = 65537$. The sum, product, and relative order of period pairs must then be computed (one feels sorry for the computerless Hermes!). The first line of the required table of information is

$$\eta_{0,2} + \eta_{1,2} = -1, \qquad \eta_{0,2} \cdot \eta_{1,2} = -2^{14}, \qquad \eta_{0,2} > \eta_{1,2}. \tag{17}$$

The product formula derives from the fact that the exponent sums evenly divide into quadratic residues and nonresidues. Drawing the Carlyle circle $C_{-1,-2^{14}}$ may place impossible demands on any compass we own, but at the conceptual level there is no problem.

Successive levels in the hierarchy of quadratic equations require $2, 4, 8, \ldots$ Carlyle circles. Working upwards however, and assuming the worst case possibility that the product formulas and sums involve all distinct periods, the largest possible number of Carlyle circles required is $1, 2, 6, 30, 270, 4590, \ldots$. Actually the 4590 is spurious, since there are only 2^{10} periods with 2^6 terms and so these can all be computed with $2^9 = 512$ Carlyle circles. Even though we know little about the details of the construction, we can conclude that the construction we seek requires $1 + 2 + 4 + 8 + \cdots + 512 + 270 + 30 + 6 + 2 + 1 = 1332$ or fewer Carlyle circles.

REFERENCES

1. R. C. Archibald, The history of the construction of the regular polygon of seventeen sides, *Bull. Amer. Math. Soc.*, 22 (1916) 239–246.
2. _____, Gauss and the regular polygon of seventeen sides, *Amer. Math. Monthly*, 27 (1920) 323–326.
3. Wayne Bishop, How to construct a regular polygon, *Amer. Math Monthly*, 85 (1978) 186–188.
4. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, 1961.
5. Howard Eves, An Introduction to the History of Mathematics, fourth edition, Holt, Rinehart, and Winston, New York, 1976.
6. C. F. Gauss, Disquisitiones Arithmeticae, Leipzig 1801, translated from the Latin by A. A. Clarke, Yale University Press, 1965.
7. Tord Hall, Carl Friedrich Gauss: A Biography, MIT Press, 1970.
8. Wilfred Keller, Factors of Fermat numbers and large primes of the form $k \times 2^n + 1$, *Math. Comp.* 41 (1983) 661–673.
9. Felix Klein, Famous Problems of Elementary Geometry, Dover, 1956 (original German edition in 1895).
10. Hans Rademacher, Lectures on Elementary Number Theory. Blaisdell, New York, 1964.
11. H. W. Richmond, *Quart. J. Math.*, 26 (1893) 296–297.

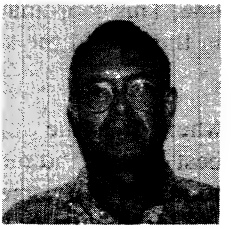
12. L. Linn Smith, A construction of the regular polygon of seventeen sides, *Amer. Math. Monthly*, 27 (1920) 322–323.
13. Ian Stewart, Gauss, *Sci. Amer.*, 237 (1977) 122–131.
14. Heinrich Tietze, Famous Problems of Mathematics, Graylock Press, New York, 1965 (original German edition in 1949).
15. P. L. Wantzel, *Journal de Mathematiques pures et appliquées*, 2 (1837) 366–372.
16. R. C. Yates, Geometrical Tools, Educational Publishers, St. Louis, 1949.
17. Jeff Young and Duncan A. Buell, The twentieth Fermat number is composite, *Math. Comp.*, 50 (1988) 261–263.

Notice to authors: Recently the MONTHLY has not been processing papers because of a very high backlog. This situation has now eased somewhat and we are once again considering new manuscripts for possible publication. Thank you for your patience. —The Editors.

Part Metric and Hyperbolic Metric

H. S. BEAR*, *University of Hawaii*

H. S. BEAR received his Ph.D. at U.C. Berkeley in 1957, under the direction of J. L. Kelley. He taught at the University of Washington, U.C. Santa Barbara, and New Mexico State University before moving to the University of Hawaii in 1969.



1. Introduction. This paper is a discussion of three metrics which have arisen in different contexts over a period of many years. One of them is the hyperbolic metric in the unit disc of the complex plane. This metric is so named because it is the metric of Poincaré's celebrated model for hyperbolic geometry. The hyperbolic metric is invariant under conformal maps of the disc onto itself, and quite naturally plays a central role in much current research in conformal and quasi-conformal mapping.

The second metric, called the part metric, was introduced by the author in the study of complex function algebras. The underlying space for a complex function algebra breaks into disjoint sets, Gleason parts, on which a kind of Schwarz lemma holds. The part metric was originally defined in these parts. The concepts of part and part metric were later extended to apply to the underlying space for a linear space of real continuous functions, and finally in a purely geometric setting, to appropriate subsets, again called parts, of a general convex set.

The third metric we consider is the Cayley-Klein-Hilbert metric. Klein used this metric in his model for hyperbolic geometry, in which the points were also the points of the open unit disc in the complex plane. Hilbert extended the definition from the disc to general bounded convex sets in Euclidean space.

These three metrics are closely related. In fact, it would be more exact to say they are identical, each being simply an avatar in the appropriate setting of one general metric. It is our purpose here to sort out the connections between the three versions, and to provide some new facts and insights about each.

We start in Section 2 with an exposition of the hyperbolic metric $\rho(z, w)$ in the unit disc. Although the hyperbolic metric is a standard tool in complex analysis, it is frequently overlooked in beginning analysis courses, and elementary explanations are hard to come by. It is easy to write down a formula for $\rho(z, w)$, and not too hard to show that distances add up correctly on the circular arcs which form the lines in Poincaré's model for hyperbolic geometry. However, it is not so easy to show that ρ is a metric!

Section 3 treats the part metric d in the open unit disc. In this setting the part metric is most naturally defined in terms of the positive harmonic functions. The

*This work was supported by a grant from the Institut d'Estudis Catalans, Centre de Recerca Matemática, Bellaterra (Barcelona). The author wishes to thank Professor Castellet of the C.R.M. and the many Spanish mathematicians who were so graciously hospitable.

Poisson kernels, which play a central role in the study of harmonic functions on the disc, are extremal among the positive harmonic functions, in a sense we will make precise later. The gradient curves of the Poisson kernels turn out to be the "lines" of the Poincaré model. Moreover, the values that a Poisson kernel function takes on each of its gradient lines constitute a coordinate system on each such hyperbolic line. This coordinate system gives both the part metric and the hyperbolic metric ρ in the disc in the sense that

$$2\rho(z,w) = d(z,w) = |\log P(z) - \log P(w)|$$

where P is the appropriate Poisson kernel. The Poisson kernels are some of the best known and most widely studied functions in analysis. It is therefore satisfying to notice that some of the results of Section 3 are both new and elementary.

Our discussion in Section 3 is not the shortest route to the conclusion that $d = 2\rho$, but the route which seems to uncover the greatest number of interesting relationships. The part metric is, unlike the hyperbolic metric, obviously a metric, so the $d = 2\rho$ equality obviates the need for a lot of geometric structure to show that ρ is a metric. Moreover, d is defined in very general plane domains, and is obviously conformally invariant. It provides, therefore, a natural generalization of ρ to these more general settings.

In Section 5 we discuss the Klein metric in the unit disc, and Hilbert's extension to general bounded open convex sets in $\mathbb{R}^n$. We indicate how this metric is very closely related to the part metric for convex sets as studied in [2], [3], [1].

Finally, in Section 6 we pose a few questions on possible extensions of these ideas.

2. The hyperbolic metric. It is most convenient, and perhaps essential, to introduce the hyperbolic metric in the context of hyperbolic geometry. Our approach will be via the Birkhoff axioms [**4**] for plane geometry (Euclidean or hyperbolic). There is an excellent comparison of the Birkhoff and Hilbert systems in Moise [**10**, p. 113]. (See also Mac Lane [**8**] for a somewhat different axiom-system in the Birkhoff vein.)

The Birkhoff axiomatic structure for plane geometry consists of a set of points, subsets of points called lines, a distance function $D(P,Q)$ on pairs of points, and a measure of angles. The distance function is assumed to have all the usual properties of a metric ($D(P,Q) \geq 0$, $D(P,Q) = D(Q,P)$, $D(P,Q) = 0$ iff $P = Q$), except that the triangle inequality is not assumed. In addition, it is assumed that each line is equipped with a coordinate system which gives the metric. That is, for each line l there is a one-to-one function f_l on l onto $(-\infty, \infty)$ such that whenever P, Q lie on l, $D(P,Q) = |f_l(P) - f_l(Q)|$. The only connection between distances on two different lines is given by the side-angle-side axiom for congruence of triangles. The triangle inequality, that $D(P,Q) + D(Q,R) > D(P,R)$ unless P, Q, R lie on a line with Q between P and R, is a theorem. The triangle inequality depends essentially on the side-angle-side axiom, but is independent of the parallel axiom, and therefore holds equally well in Euclidean and hyperbolic geometry.

In the Poincaré model for hyperbolic geometry, the points are the points of the open unit disc D in the complex plane. Thus the points can be identified with the complex numbers z with $|z| < 1$. The "lines" in this model are the intersections with D of circles or lines perpendicular to the unit circle Γ. A few pictures will

convince the reader that two points determine a unique "line"; i.e., for each pair z, w there is exactly one circular arc which goes through z and w and is perpendicular to Γ. If z and w lie on a diameter of D, then the diameter is the "circular arc" through z and w. For each pair α, β of distinct points of Γ there is a unique hyperbolic line (hereafter "h-line") which is an open circular arc from α to β, perpendicular to Γ at both α and β. We will use (α, β) to denote this h-line, and emphasize that α and β are not points of the line or points of the model.

Let us recall some basic properties of the Möbius transformations (fractional linear transformations) which are functions of the form

$$T(z) = \frac{az + b}{cz + d}, \qquad ad - bc \neq 0. \tag{1}$$

The transformations (1) are one-to-one conformal maps of the complex plane plus the point at infinity onto itself. Möbius transformations form a group under composition, and each transformation takes every circle or line onto a circle or line. Since there are just three essential constants in (1), any Möbius transformation Γ is determined by specifying any three values: $T(z_i) = w_i$, $i = 1, 2, 3$.

The Möbius transformations of the form

$$T(z) = \mu \cdot \frac{z - z_0}{1 - \bar{z}_0 z}, \qquad |\mu| = 1, \quad |z_0| < 1, \tag{2}$$

map the unit circle Γ onto itself, and the disc D onto itself. We will call the functions (2) disc-maps. The disc maps form a subgroup of the Möbius transformations.

Notice that any disc map (2) maps any h-line (α, β) onto some other h-line $(T(\alpha), T(\beta))$. This is because T maps Γ onto Γ, and the circular arc (α, β) onto some other circular arc in D. The image arc is also perpendicular to Γ, because T is a conformal mapping, and so the image arc is again an h-line.

We will use the cross ratio, defined next, to put a coordinate system on each h-line, and these coordinate systems will give the hyperbolic metric in D. For any four complex numbers z_0, z_1, z_2, z_3, with z_1, z_2, z_3 distinct, define the cross ratio as follows:

$$(z_0, z_1, z_2, z_3) = \frac{(z_2 - z_0)(z_3 - z_1)}{(z_1 - z_0)(z_3 - z_2)}. \tag{3}$$

A geometric interpretation of the cross ratio will appear in Section 5. Notice that interchanging z_1 and z_2, or interchanging z_0 and z_3 gives the reciprocal cross-ratio:

$$\begin{aligned} (z_0, z_2, z_1, z_3) &= (z_0, z_1, z_2, z_3)^{-1}, \\ (z_3, z_1, z_2, z_0) &= (z_0, z_1, z_2, z_3)^{-1}. \end{aligned} \tag{4}$$

For distinct z_1, z_2, z_3, we define a Möbius transformation T by

$$T(z) = (z, z_1, z_2, z_3) = \left[\frac{z - z_2}{z - z_1}\right] \cdot \left[\frac{z_3 - z_1}{z_3 - z_2}\right]. \tag{5}$$

T is the Möbius transformation which sends z_1, z_2, z_3, respectively, to $\infty, 0, 1$.

The fact that the hyperbolic metric is invariant under disc maps will depend on the fact that cross ratios are invariant under disc maps, and indeed under all

Möbius transformations. To see this, let z_1, z_2, z_3 be any three distinct points and let $T(z) = (z, z_1, z_2, z_3)$ be the Möbius transformation which sends z_1, z_2, z_3 to $\infty, 0, 1$ respectively. Let R be any Möbius transformation, and let

$$S(z) = (z, R(z_1), R(z_2), R(z_3)). \tag{6}$$

Then S is of course the Möbius transformation which sends $R(z_1), R(z_2), R(z_3)$ to $\infty, 0, 1$ respectively. Hence $S^{-1} \circ T$ sends z_1, z_2, z_3 to $R(z_1), R(z_2), R(z_3)$, respectively. Since three values determine a Möbius transformation, $S^{-1} \circ T = R$, or $T = S \circ R$. Thus

$$T(z_0) = (z_0, z_1, z_2, z_3) = S(R(z_0)) = (R(z_0), R(z_1), R(z_2), R(z_3)), \tag{7}$$

and the cross ratio is invariant under any Möbius transformation R.

The argument above also shows that any three points z_1, z_2, z_3 can be mapped onto any three points w_1, w_2, w_3 by some Möbius transformation; specifically, if S sends z_1, z_2, z_3 to $\infty, 0, 1$ and T sends w_1, w_2, w_3 to $\infty, 0, 1$, then $T^{-1} \circ S$ sends z_i to w_i. Since all Möbius transformations map circles and lines onto circles or lines, and circles (lines) are determined by three points, it follows that any circle or line can be mapped onto any circle or line by a Möbius transformation.

The cross ratio (z_0, z_1, z_2, z_3) is real if all z_i lie on the real line. Hence (z_0, z_1, z_2, z_3) is real if and only if all z_i lie on some circle or line, since any such circle or line could be mapped conformally onto the real line, and the cross ratio of the four points would not change. In particular, if z, w are points of the disc and lie on the h-line (α, β), then (α, z, w, β) is real. Moreover, (α, z, w, β) is positive since a mapping of the circle containing (α, β) onto the real line would send z and w to points between the images of α and β.

Now we introduce a coordinate system on the h-line $(-1, 1)$. Since any h-line can be disc-mapped onto $(-1, 1)$, this will provide a coordinate system for every h-line. For $-1 < r < 1$, let

$$f(r) = \frac{1}{2}\log\frac{1+r}{1-r}. \tag{8}$$

Clearly f is a one-to-one map on $(-1, 1)$ onto $(-\infty, \infty)$. According to this coordinate system, the hyperbolic distance ρ between 0 and r is

$$\rho(0, r) = |f(r)| = \frac{1}{2}\left|\log\frac{1+r}{1-r}\right|, \tag{9}$$

and the h-distance between points r_1 and r_2 is

$$\begin{aligned}
\rho(r_1, r_2) &= |f(r_2) - f(r_1)| \\
&= \frac{1}{2}\left[\log\frac{1+r_2}{1-r_2} - \log\frac{1+r_1}{1-r_1}\right] \\
&= \frac{1}{2}\left|\log\left[\frac{r_2-(-1)}{r_1-(-1)} \cdot \frac{1-r_1}{1-r_2}\right]\right| \\
&= \frac{1}{2}|\log(-1, r_1, r_2, 1)|,
\end{aligned} \tag{10}$$

where $(-1, r_1, r_2, 1)$ is the cross-ratio defined in (3). The factor $\frac{1}{2}$ in (8), (9), and

(10) is there for the following reason. The hyperbolic metric can also be characterized as the minimum length of the curves γ from z to w, where the length of γ is given by the line integral

$$\text{length}(\gamma) = \int_\gamma \frac{|dz|}{1 - |z|^2}. \tag{11}$$

For $z = 0$, $w = r > 0$, the γ of minimum length is the segment $[0, r]$ and the integral (11) gives (9) with the factor $\frac{1}{2}$ as indicated. The factor $\frac{1}{2}$ also gives the normalizing condition $\lim_{z \to 0} \rho(0, z)/|z| = 1$.

If (α, β) is any h-line, there are many disc maps T which map (α, β) onto $(-1, 1)$. For any such T, the function

$$f_{(\alpha,\beta)}(z) = f(T(z)) \tag{12}$$

puts a coordinate system on (α, β). For any points $z, w \in (\alpha, \beta)$ we then have

$$\begin{aligned} \rho(z,w) &= |f_{(\alpha,\beta)}(z) - f_{(\alpha,\beta)}(w)| \\ &= |f(T(z)) - f(T(w))| \\ &= \frac{1}{2}|\log(-1, T(z), T(w), 1)| \\ &= \frac{1}{2}|\log(\alpha, z, w, \beta)|. \end{aligned} \tag{13}$$

The last equality holds if $T(\alpha) = -1$, $T(\beta) = 1$, no matter which disc map T is used, because cross ratio is invariant under disc maps. From (4) we see that

$$(\beta, z, w, \alpha) = (\alpha, z, w, \beta)^{-1}, \tag{14}$$

so (12) still holds if T maps α to 1 and β to -1 instead of the other way around. Notice that the method above of putting a coordinate system on each h-line automatically gives a distance function which is invariant under disc maps. Notice also that while different disc maps T put different coordinate systems on (α, β), they all yield the same distance formula.

To summarize, we define the hyperbolic metric ρ on D as follows. For any points z, w, there is a unique h-line (α, β) determined by z and w. The cross ratio (α, z, w, β) is real and positive for $z, w \in (\alpha, \beta)$, and

$$\rho(z,w) = \tfrac{1}{2}|\log(\alpha, z, w, \beta)|. \tag{15}$$

Interchanging z and w, or α and β, gives the reciprocal of the cross ratio, so the absolute value of the logarithm does not change. The hyperbolic metric (15) is invariant under disc maps, since the cross ratio is, and the distances add up on h-lines ($\rho(z_1, z_2) + \rho(z_2, z_3) = \rho(z_1, z_3)$ when z_2 is between z_1 and z_3) because distances on each line are given by a coordinate system.

As a last but very important observation we note that the triangle inequality holds for ρ; namely, $\rho(z_1, z_2) + \rho(z_2, z_3) > \rho(z_1, z_3)$ unless z_1, z_2, z_3 are h-colinear with z_2 between z_1 and z_3. We have not shown this, and it is not obvious. The triangle inequality is a theorem of both Euclidean and hyperbolic geometry which depends in an essential way on the side-angle-side axiom for congruence of triangles. Thus the triangle inequality for the hyperbolic metric in the disc depends

on the non-trivial verification of the side-angle-side axiom for hyperbolic triangles in the Poincaré model. We will give a simpler proof in Section 4 using the equality between the part metric and the hyperbolic metric.

3. The part metric in D. If z, w are any two points of D, then there exists a constant $a > 0$ depending only on z and w such that for every positive harmonic function u on D

$$\frac{1}{a} < u(z)/u(w) < a. \tag{1}$$

Equation (1) is called Harnack's inequality, and holds for all positive harmonic functions and points z, w in any plane domain, or indeed any domain in $\mathbb{R}^n$.

Suppose instead of positive harmonic functions on D we consider the positive functions in any linear space B of (continuous) functions on any set X. The space B could consist of the real parts of the complex functions in some function algebra, for example. The existence of a constant a so that (1) holds for all positive functions in B defines an equivalence relation, $z \sim w$, on the points of X. The equivalence classes of this relation are the Gleason parts of X with respect to B, and these parts were the original setting for the part metric [2].

To return to the disc D, we define the part metric $d(z,w)$ for z, w in D as follows:

$$\begin{aligned} d(z,w) &= \sup_{u>0} |\log u(z)/u(w)| \\ &= \sup_{u>0} |\log u(z) - \log u(w)|. \end{aligned} \tag{2}$$

The supremum is over all positive harmonic functions on D, and is finite because of the Harnack inequality (1). From the last equality in (2) it is clear that d is a metric. It is also immediate from the definition that d is invariant under disc maps T, for $v = u \circ T$ is a positive harmonic function on D iff u is, so

$$\begin{aligned} d(T(z),T(w)) &= \sup_{u>0} |\log u(T(z)) - \log u(T(w))| \\ &= \sup_{v>0} |\log v(z) - \log v(w)| \\ &= d(z,w). \end{aligned} \tag{3}$$

Let us denote by H^+ the positive harmonic functions on D, and by $H^+(z_0)$ the positive harmonic functions u such that $u(z_0) = 1$. If we give the harmonic functions on D the topology of uniform convergence on compact subsets of D, then $H^+(z_0)$ is a compact set for each z_0. Since $H^+(z_0)$ is also obviously convex, the Krein-Milman theorem [**11**, p. 207] guarantees that $H^+(z_0)$ has extreme points, and is the closed convex hull of those extreme points. The theorem of Herglotz [**12**, Th. 11.19] states that $u \in H^+(0)$ iff there is a unique positive probability measure μ on Γ such that

$$u(z) = \int_\Gamma P_\alpha(z)\, d\mu(\alpha), \tag{4}$$

where $P_\alpha(z)$ is the Poisson kernel

$$P_\alpha(z) = (1 - |z|^2)/|\alpha - z|^2, \quad |\alpha| = 1. \tag{5}$$

The Herglotz theorem is an explicit form of the Krein-Milman theorem for the compact convex set $H^+(0)$. It puts $H^+(0)$ into a one-to-one affine correspondence with probability measures on Γ. The extreme points of the set of measures are the point masses (one at each $\alpha \in \Gamma$) so the extreme points on $H^+(0)$ are the Poisson kernels $P_\alpha(z)$, one for each $\alpha \in \Gamma$.

A positive harmonic function $Q \in H^+$ was called by R. S. Martin [9] a *minimal function* if $0 < u \leq Q$ and $u \in H^+$ imply that $u = cQ$ for some positive constant c. The minimal functions are sometimes easier to work with than the extreme points of the compact sets $H^+(z_0)$, and the following lemma makes clear the relationship between these two ideas.

LEMMA 1. *The minimal functions are exactly the positive multiples of the Poisson kernels. The extreme points of $H^+(z_0)$ are the functions $P_\alpha/P_\alpha(z_0)$; i.e., the Poisson kernels normalized to be one at z_0, or equivalently the minimal functions normalized at z_0.*

Proof. Let Q be a minimal function and let z_0 be arbitrary. We show that $Q/Q(z_0)$ is an extreme point of $H^+(z_0)$, and so in particular $Q/Q(0)$ is a Poisson kernel.

If, contrary to our assertion, $Q/Q(z_0)$ is not an extreme point of $H^+(z_0)$, then there are $u, v \in H^+(z_0)$, and $\lambda \in (0, 1)$ such that

$$\frac{Q}{Q(z_0)} = \lambda u + (1 - \lambda)v. \tag{6}$$

Clearly $Q/Q(z_0)$ is a minimal function, since Q is. Both functions on the right side of (6) are positive, so both are less than $Q/Q(z_0)$. It follows that both u and v are multiples of $Q/Q(z_0)$ and hence of Q. Thus $u = v = Q/Q(z_0)$ and $Q/Q(z_0)$ must be an extreme point of $H^+(z_0)$.

Now assume that $Q/Q(z_0)$ is an extreme point of $H^+(z_0)$, for some z_0. To show that Q is minimal, assume that $u \in H^+$ and $u < Q$. Then $Q - u$ and $Q + u$ are in H^+, and in $H^+(z_0)$ when suitably normalized. Since Q is the average of $Q - u$ and $Q + u$,

$$\begin{aligned}\frac{Q}{Q(z_0)} = {} & \frac{1}{2}\frac{(Q(z_0) - u(z_0))}{Q(z_0)}\left[\frac{Q - u}{Q(z_0) - u(z_0)}\right] \\ & + \frac{1}{2}\frac{(Q(z_0) + u(z_0))}{Q(z_0)}\left[\frac{Q + u}{Q(z_0) + u(z_0)}\right].\end{aligned} \tag{7}$$

Equation (7) displays $Q/Q(z_0)$ as a convex combination of two functions (in brackets) of $H^+(z_0)$. Since $Q/Q(z_0)$ is extreme, both of the functions in brackets equal $Q/Q(z_0)$, so u is a multiple of Q, which shows Q is minimal.

LEMMA 2. *For any $z, w \in D$ there are Poisson kernels P_α, P_β such that*

$$\max_{u \in H^+} u(z)/u(w) = P_\alpha(z)/P_\alpha(w), \tag{8}$$

$$\min_{u \in H^+} u(z)/u(w) = P_\beta(z)/P_\beta(w). \tag{9}$$

It follows that

$$d(z, w) = \left|\log P_\gamma(z)/P_\gamma(w)\right| \tag{10}$$

for some Poisson kernel P_γ ($\gamma = \alpha$ or $\gamma = \beta$).

Notice that it is not clear at this point which is larger, the maximum value of $u(z)/u(w)$, or the reciprocal of the minimum value of $u(z)/u(w)$, or how the distinction depends on z and w.

Proof. Fix z and w, and let F be the linear functional on H defined by $F(u) = u(z)$. Clearly F is continuous, and so [**11**, p. 208] assumes its maximum on the compact set $H^+(w)$ at some extreme point $P_\alpha/P_\alpha(w)$. Hence, for some Poisson kernel P_α,

$$\begin{aligned}\sup\{\log u(z)/u(w) : u \in H^+\} &= \sup\{\log u(z) : u \in H^+(w)\} \\ &= \sup\{\log F(u) : u \in H^+(w)\} \\ &= \log P_\alpha(z)/P_\alpha(w). \end{aligned} \tag{11}$$

Similarly, for some P_β

$$\begin{aligned}\inf\{-\log u(z)/u(w) : u \in H^+\} &= \inf\{-\log u(z) : u \in H^+(w)\} \\ &= \min\{-\log F(u) : u \in H^+(w)\} \\ &= -\log P_\beta(z)/P_\beta(w) \\ &= \log P_\beta(w)/P_\beta(z). \end{aligned} \tag{12}$$

Let us notice the connection between the Poisson kernel $P_\alpha(z)$, $|\alpha| = 1$, and the Möbius transformation

$$T_\alpha(z) = \frac{\alpha + z}{\alpha - z} = \frac{(\alpha + z)(\bar{\alpha} - \bar{z})}{(\alpha - z)(\bar{\alpha} - \bar{z})} = \frac{1 - |z|^2}{|\alpha - z|^2} + \frac{\bar{\alpha}z - \alpha\bar{z}}{|\alpha - z|^2}.$$

Since $\bar{\alpha}z - \alpha\bar{z}$ is pure imaginary, we have

$$P_\alpha(z) = \operatorname{Re}\left(\frac{\alpha + z}{\alpha - z}\right).$$

If $|z| = 1$, $z \neq \alpha$, then $\operatorname{Re} T_\alpha(z) = 0$. Hence T_α maps the unit circle onto the imaginary axis. Since $T_\alpha(0) = 1$, T_α maps the disc onto the right half-plane. The level curves of the Poisson kernel P_α correspond to the lines $x =$ constant under the mapping T_α. It follows that the level curves of P_α are circles in D, tangent to Γ at α.

The gradient lines for P_α are of course perpendicular to the level curves. The gradient lines therefore correspond under T_α to the rays $y =$ constant, $x > 0$. These gradients are therefore also circular arcs in D, from α to all other points of Γ. In other words, the gradient lines of P_α are exactly all hyperbolic lines ending at α, and $P_\alpha(z)$ goes from 0 to $+\infty$ along each such h-line. This suggests that if $z, w \in (\alpha, \beta)$, with $P_\alpha(z) > P_\alpha(w)$ (and $P_\beta(w) > P_\beta(z)$), then the part metric distance might be given by $\log P_\alpha(z)/P_\alpha(w)$ or by $\log P_\beta(w)/P_\beta(z)$. The next results show that this is indeed the case.

LEMMA 3. *If T is any disc map, then Q is a minimal function iff $Q \circ T$ is a minimal function.*

Proof. For all u, $u \in H^+$ iff $u \circ T \in H^+$, and $0 < u < Q$ iff $0 < u \circ T < Q \circ T$. If Q is minimal, and $0 < v < Q \circ T$, then $v = u \circ T$ for some u, with $0 < u < Q$.

Hence $u = cQ$, $v = cQ \circ T$, and $Q \circ T$ is minimal if Q is. Conversely, if $Q \circ T$ is minimal, then $Q \circ T \circ T^{-1} = Q$ is minimal.

THEOREM 4. *If α, β are points of the unit circle Γ and P_α, P_β are the corresponding Poisson kernels* (5), *there is a constant $C_{\alpha\beta}$ so that for all $z \in (\alpha, \beta)$*

$$P_\alpha(z)P_\beta(z) = C_{\alpha\beta}. \tag{13}$$

It follows that for all $z, w \in (\alpha, \beta)$,

$$P_\alpha(z)/P_\alpha(w) = P_\beta(w)/P_\beta(z). \tag{14}$$

Notice that from (14) it follows that if $z, w \in (\alpha, \beta)$ and if the part distance $d(z, w) = \log P_\alpha(z)/P_\alpha(w)$, then

$$d(z, w) = \log P_\alpha(z)/P_\alpha(w) = \log P_\beta(w)/P_\beta(z). \tag{15}$$

It remains to be shown that $d(z, w)$ is indeed given as above.

Proof. Fix $z_0 \in (\alpha, \beta)$ and let T be a disc map which sends z_0 to 0 and β to 1. ($T(z) = \mu(z - z_0)/(1 - \bar{z}_0 z)$ sends z_0 to 0 and Γ onto Γ. For the right choice of μ, $|\mu| = 1$, β will go to 1.) We have seen that disc maps send hyperbolic lines onto hyperbolic lines, so T maps (α, β) onto the hyperbolic line through 0, ending at 1. Hence T maps (α, β) onto the diameter $(-1, 1)$. Since P_1 (i.e., P_α with $\alpha = 1$) is a minimal function, by Lemma 3, $P_1 \circ T$ is a minimal function. Since $P_1 \circ T$ has its pole at β, $P_1 \circ T = cP_\beta$ for some constant c. Moreover, $P_1 \circ T(z_0) = P_1(0) = 1$, so $c = 1/P_\beta(z_0)$, and

$$P_\beta = P_\beta(z_0)(P_1 \circ T). \tag{16}$$

The same argument applied to $-T$ shows that

$$P_\alpha = P_\alpha(z_0)(P_1 \circ (-T)). \tag{17}$$

Let $z \in (\alpha, \beta)$, and $T(z) = r \in (-1, 1)$. Then

$$P_\beta(z) = P_\beta(z_0)P_1(r) = P_\beta(z_0)[(1 + r)/(1 - r)], \tag{18}$$

and

$$P_\alpha(z) = P_\alpha(z_0)P_1(-r) = P_\alpha(z_0)[(1 - r)/(1 + r)]. \tag{19}$$

Hence for any $z \in (\alpha, \beta)$

$$P_\alpha(z)P_\beta(z) = P_\alpha(z_0)P_\beta(z_0). \tag{20}$$

Since z_0 was arbitrary, we have the result.

COROLLARY. *For any $\alpha, \beta \in \Gamma$ and any $z \in (\alpha, \beta)$,*

$$P_\alpha(z)P_\beta(z) = \frac{(1 - |z|^2)^2}{|\alpha - z|^2|\beta - z|^2} = C_{\alpha\beta} = \frac{4}{|\alpha - \beta|^2}. \tag{21}$$

Proof. As $z \to \beta$ on (α, β), z approaches β radially in the limit, so $|\beta - z|/(1 - |z|) \to 1$, and

$$\lim_{z \to \beta} \frac{(1 - |z|^2)^2}{|\alpha - z|^2|\beta - z|^2} = \lim_{z \to \beta} \frac{(1 - |z|)^2}{|\alpha - z|^2} = \frac{4}{|\alpha - \beta|^2}. \tag{22}$$

The next lemma and theorem show that if $z, w \in (\alpha, \beta)$, in the order α, z, w, β, then $d(z, w)$ is indeed given by $\log P_\alpha(z)/P_\alpha(w) = \log P_\beta(w)/P_\beta(z)$.

LEMMA 5. *If $z, w \in (\alpha, \beta)$, and T is a disc map which sends α, z, w, β to $1, 0, r, 1$, with $r > 0$, then*

$$d(z, w) = \log(1 + r)/(1 - r) = \log P_1(r)/P_1(0).$$

Proof. Since d is invariant under disc maps, $d(z, w) = d(0, r)$. By definition, and Lemma 2, and Theorem 4, we can write

$$\begin{aligned} d(0, r) &= \max_{\gamma \in \Gamma} \log P_\gamma(r)/P_\gamma(0) \\ &= \max_{\gamma \in \Gamma} \log P_\gamma(r) \\ &= \max_{\gamma \in \Gamma} \log(1 - r^2)/|\gamma - r|^2. \end{aligned} \tag{23}$$

Clearly $|\gamma - r|$ is minimum when $\gamma = 1$, so

$$\begin{aligned} d(0, r) &= \log(1 - r^2)/(1 - r)^2 \\ &= \log(1 + r)/(1 - r) \\ &= \log P_1(r). \end{aligned} \tag{24}$$

THEOREM 6. *If $z, w \in (\alpha, \beta)$, in the order α, z, w, β, then*

$$d(z, w) = \log P_\alpha(z)/P_\alpha(w) = \log P_\beta(w)/P_\beta(z). \tag{25}$$

Proof. The two ratios are the same by Theorem 4. There is some disc map T which sends α, z, w, β respectively to $-1, 0, r, 1$ with $r > 0$. Recalling that $P_1 \circ T = cP_\beta$, we have

$$\begin{aligned} d(z, w) &= d(0, r) \\ &= \log P_1(r)/P_1(0) \\ &= \log P_1(T(w))/P_1(T(z)) \\ &= \log P_\beta(w)/P_\beta(z). \end{aligned} \tag{26}$$

COROLLARY 1. *For all $z, w \in D$,*

$$d(z, w) = 2\rho(z, w). \tag{27}$$

Proof. By the definition (2-8), if $r > 0$, then

$$\rho(0, r) = \tfrac{1}{2}\log(1 + r)/(1 - r).$$

By the result above (Lemma 5), $d(0, r) = \log(1 + r)/(1 - r)$. For any two points z and w there is some disc map T which sends z to 0 and w to some $r > 0$. Since both d and ρ are invariant under disc maps, the result follows from the fact that $d(0, r) = 2\rho(0, r)$.

COROLLARY 2. *If $z, w \in (\alpha, \beta)$ in the order α, z, w, β. Then*

$$d(z, w) = \log(\alpha, z, w, \beta). \tag{28}$$

Proof. This follows from Corollary 1 and (10) of Section 2.

COROLLARY 3. *If* $z, z_0 \in (\alpha, \beta)$ *with* $P_\beta(z_0) = 1$, *then*

$$P_\beta(z) = (\alpha, z_0, z, \beta). \tag{29}$$

Proof. If $P_\beta(z_0) = 1$, then using (28)

$$\begin{aligned} d(z_0, z) &= |\log P_\beta(z)| \\ &= |\log(\alpha, z_0, z, \beta)|. \end{aligned} \tag{30}$$

Since $P_\beta(z) > 1$ and $(\alpha, z_0, z, \beta) > 1$ iff z is closer to β than z_0, (29) follows.

Now we summarize our result. The function $g(r) = (1 + r)/(1 - r)$ is a one-one map of $(-1, 1)$ onto $(0, \infty)$ and hence $f(r) = \frac{1}{2}\log(1 + r)/(1 - r)$ is a one-one map of $(-1, 1)$ onto $(-\infty, \infty)$. Thus f puts a coordinate system on $(-1, 1)$, and the hyperbolic distance between points r_1, r_2 of $(-1, 1)$ according to this coordinate system is

$$\rho(r_1, r_2) = |f(r_2) - f(r_1)| = \tfrac{1}{2}\log(-1, r_1, r_2, 1). \tag{31}$$

Of course $\rho(r_1, r_2) + \rho(r_2, r_3) = \rho(r_1, r_3)$ if $-1 < r_1 < r_2 < r_3 < 1$. By mapping any h-line (α, β) onto $(-1, 1)$ with a disc map T, the coordinate system can be carried to (α, β) giving the distance

$$\rho(z, w) = \rho(T(z), T(w)) = \tfrac{1}{2}|\log(-1, T(z), T(w), 1)| \tag{32}$$

for points $z, w \in (\alpha, \beta)$. Distances add up correctly on each h-line (α, β) because they do on $(-1, 1)$. Since cross ratios are invariant under disc maps, the distances on (α, β) can also be given by

$$\rho(z, w) = \tfrac{1}{2}|\log(\alpha, z, w, \beta)|. \tag{33}$$

The triangle inequality for ρ is not obvious, and follows as a theorem in either Euclidean or hyperbolic geometry once the side-angle-side axiom for triangle congruence is verified.

The part metric d is defined by

$$d(z, w) = \sup\{|\log u(z)/u(w)|: u \in H^+\}. \tag{34}$$

This definition makes sense on any plane domain on which there are non-constant positive harmonic functions. (It is enough to assume the complement of the domain has at least two points.) More generally, the definition (34) makes sense on the Gleason parts of any function space. For z, w, in the disc, the sup in (34) is attained for both $u = P_\alpha$ and for $u = P_\beta$, where $z, w \in (\alpha, \beta)$, and

$$d(z, w) = |\log P_\alpha(z) - \log P_\alpha(w)| = |\log P_\beta(z) - \log P_\beta(w)|. \tag{35}$$

The part metric is invariant under conformal maps for any domain, and in the disc, $d = 2\rho$.

4. The Schwarz-Pick lemma and the Riemann mapping theorem. The fact that the part metric and the hyperbolic metric coincide in the disc makes possible a very simple proof of the Schwarz-Pick lemma, and provides a more intuitive variant of the proof of the Riemann mapping theorem. We start with a simple proof of the fact that equality holds in the triangle inequality only for colinear points.

LEMMA 1. *If $z, \zeta, w \in D$ and $d(z,\zeta) + d(\zeta,w) = d(z,w)$, then ζ lies on the hyperbolic interval $[z,w]$.*

Proof. It suffices to consider the case $z = 0$ and $w = r > 0$ since any z and w can be disc-mapped into these points with distances preserved. So assume that

$$d(0,\zeta) + d(\zeta,r) = d(0,r) \tag{1}$$

for some $r > 0$ and some $\zeta \in D$. Clearly $d(0,\zeta) = d(0,|\zeta|)$ since rotation is a disc map and preserves distances. It follows that $0 \le |\zeta| \le r$ since $d(\zeta,r) \ge 0$. Since $d(0,\zeta) = d(0,|\zeta|)$, it follows that

$$d(\zeta,r) = d(|\zeta|,r). \tag{2}$$

We know that

$$\begin{aligned} d(|\zeta|,r) &= \log P_1(r)/P_1(|\zeta|) \\ &= \log \frac{1-r^2}{(1-r)^2}\,\frac{(1-|\zeta|)^2}{1-|\zeta|^2}, \end{aligned} \tag{3}$$

and that

$$\begin{aligned} d(\zeta,r) &\ge \log P_1(r)/P_1(\zeta) \\ &= \log \frac{1-r^2}{(1-r)^2}\,\frac{|1-\zeta|^2}{1-|\zeta|^2}. \end{aligned} \tag{4}$$

The right sides of (3) and (4) differ only by the factor $(1-|\zeta|)^2$ in (3) and the factor $|1-\zeta|^2$ in (4). Clearly $|1-\zeta| > 1-|\zeta|$ unless ζ is real, so equality in (1) implies that $\zeta \in [0,r]$.

THEOREM 2. (Schwarz-Pick lemma [**6**, p. 41]). *If f is an analytic map of D into D and f preserves the hyperbolic distance between any two points, then f is a disc map and preserves all distances.*

Proof. It is immediate from the definition that any analytic map decreases ($\le$) all distances:

$$\begin{aligned} d(f(z),f(w)) &= \sup_{u>0} |\log u(f(z)) - \log u(f(w))| \\ &\le \sup_{v>0} |\log v(z) - \log v(w)|, \end{aligned} \tag{5}$$

since every $u \circ f$ is a harmonic v, but not vice-versa unless f is one-to-one onto D. Suppose that $d(f(z),f(w)) = d(z,w)$ for some pair z,w, and let ζ be any point of the hyperbolic segment $[z,w]$. Then

$$\begin{aligned} d(z,w) &= d(z,\zeta) + d(\zeta,w) \\ &\ge d(f(z),f(\zeta)) + d(f(\zeta),f(w)) \\ &\ge d(f(z),f(w)) = d(z,w). \end{aligned} \tag{6}$$

It follows that equality holds throughout in (6), and that $f(\zeta)$ lies on the hyperbolic segment $[f(z),f(w)]$ and moreover that $f(\zeta)$ is the same distance from $f(z)$ that ζ is from z. If T is a disc map which sends z into $f(z)$ and the ray from z toward w onto the ray from $f(z)$ toward $f(w)$, then T also preserves distances along the

line, so $T(\zeta) = f(\zeta)$ for all $\zeta \in [z, w]$. Consequently $f \equiv T$, and f preserves all distances.

We now outline a proof of the Riemann mapping theorem which provides a more geometric intuition than the usual one. First one shows as usual that if G is a simply connected proper subdomain of $\mathbb{C}$ there is a one-to-one map f_0 of G into the disc D. If $z_0 \notin G$, then $h(z) = \sqrt{z - z_0}$ opens the plane up so that some disc $D(a, r) = \{z: |z - a| \leq r\}$ is missing from $h[G]$. Since $|h(z) - a| \geq r$ for $z \in G$, $f_0(z) = r[h(z) - a]^{-1}$ has modulus less than one on G. Now consider the set S of all one-to-one analytic maps f on G into D which spread every pair, z, w farther apart in D than f_0 does; i.e., all f so that

$$d(f(z), f(w)) \geq d(f_0(z), f_0(w)) \tag{7}$$

for all $z, w \in G$. The set S is compact in the topology of uniform convergence on compact subsets of G. This is because S is a normal family (a precompact set), and a limit function is clearly also one-to-one, since it spreads all pairs z, w at least as much as f_0 does.

For any fixed pair $z_0, w_0 \in G$, the function $d(f(z_0), f(w_0))$ is a continuous function of f, for $f \in S$. Hence there is some $g \in S$ such that $d(g(z_0), g(w_0))$ is maximum. If $g[G]$ is not all of D, let $a \notin g[G]$, and let $T_a(z) = (z - a)/(1 - \bar{a}z)$ map a onto 0. Then $T_a \circ g$ maps z_0 and w_0 maximally far apart, and 0 is not in the range of $T_a \circ g$. There is a square root branch R which is one-to-one on the simply connected range of $T_a \circ g$. Since the squaring function on D to D is not one-to-one, it strictly decreases all distances. Hence R strictly increases all distances for points in its domain. Hence $R \circ T_a \circ g$ is a one-to-one function on G to D which sends z_0 and w_0 farther apart than g does. The contradiction shows that g does indeed map g onto D if g maps any pair z_0 and w_0 maximally far apart.

5. The Klein-Hilbert part metric in a convex set. In Klein's model for hyperbolic geometry, the points are the points of the open unit disc D, and the lines are the chords in D. If z, w are points of D and α and β are the endpoints of the chord through z and w, then the Klein metric is given by $\delta(z, w) = |\log(\alpha, z, w, \beta)|$, where (α, z, w, β) again denotes the cross-ratio. Hilbert later observed that this metric works for points of an arbitrary bounded open convex set in $\mathbb{R}^n$ (see [**5**, p. 105]). We show below that the part metric as extended to convex sets in [**2**], [**3**], [**1**] is essentially the same as the Klein metric.

If K is any convex set we define an equivalence relation $x \sim y$ for points $z, y \in K$ by agreeing that $x \sim y$ iff the segment $[x, y]$ extends beyond x and y in K. That is, $x \sim y$ iff x and y are both interior points of some segment $[x', y'] \subset K$. A simple two-dimensional picture shows that if the segment $[x, y]$ extends both ways to a segment $[x', y']$ in K, and $[y, z]$ extends to a segment $[y'', z']$ in K, then $[x, z]$ extends to a segment $[x'', z'']$ where x'' lies on $[x', y'']$ and z'' lies on $[y', z']$. Hence $x \sim y$ and $y \sim z$ imply $x \sim z$, and $\sim$ is an equivalence relation.

The equivalence classes of the above relation are called the parts of K. For example, if K is a closed tetrahedron, the four extreme points are singleton parts, the six open edges are parts, the four open faces are parts, and the interior is one part. If K is an open convex set in $\mathbb{R}^n$, the K has just one part; i.e., $x \sim y$ for all $x, y \in K$.

This concept of part for a convex set becomes a generalization of the original idea of part for a function algebra or function space in the following way. Let B be

a linear space of continuous real functions u on a compact space X. The functions u are represented as linear functionals on the unit ball of the adjoint space B^*. This ball is a convex set K, and compact in the appropriate topology. There is a constant a so that $1/a < u(x)/u(y) < a$ for all positive u in B, if and only if the segment $[x, y]$ extends in K; i.e., the two concepts of part coincide. Here we are using x and y to represent both points of X and the corresponding evaluation functionals in B^*. Thus the parts of X are just the intersections of X with the parts of the unit ball K of B^*. (The details are in [2].)

There is not necessarily a distance function in a general convex set, so the segments in a general K do not have a length a priori. Nevertheless, the ratio of lengths of segments on the same line makes sense, since we can define $\overline{ay}/\overline{ax}$ to be r if $y = a + r(x - a)$. The ratio $\overline{ay}/\overline{ax}$ can also be thought of as the maximum of the ratios $u(y)/u(x)$ as u ranges over positive linear functions on the segment $[a, b]$. A picture of a positive linear function on a real interval $[a, b]$ with $a < x < y < b$, shows that $u(y)/u(x)$ is maximum for $u \geq 0$ when $u(a) = 0$, in which case $u(y)/u(x)$ is the same as $\overline{ay}/\overline{ax}$.

The Klein metric in a convex set is

$$\delta(x, y) = \log\frac{\overline{ay}}{\overline{ax}} + \log\frac{\overline{xb}}{\overline{yb}} = \log(a, x, y, b). \tag{1}$$

The part metric would be

$$d(x, y) = \sup\left\{\left|\log\frac{u(x)}{u(y)}\right| : u > 0, u \text{ linear}\right\}, \tag{2}$$

which would be the larger of $\log(\overline{ay}/\overline{ax})$ and $\log(\overline{xb}/\overline{yb})$. Thus the Klein metric looks at how close the points are to the boundary on both sides of segment $[x, y]$, whereas the part metric looks only at the closest boundary point. Because of this one-sided aspect of the part metric, the part metric is defined in any convex set which contains no whole line, whereas the Klein metric is defined only for convex sets which intersect every line in a bounded interval.

If one looks at the details of the proof of the triangle inequality for the part or Klein metric in a convex set [3], one sees that equality can hold ($d(x, y) + d(y, z) = d(x, z)$) for points x, y, z which are not colinear. This happens, however, only when the convex set contains line segments in its boundary. For example, if K is the interior of the unit square, $K = (0, 1) \times (0, 1)$, then $d(x, y) + d(y, z) = d(x, z)$ whenever x, y, z are points such that the line determined by x and y and the line determined by y and z both intersect ∂K only on the vertical sides. If K is strictly convex, so there are no line segments on ∂K, then equality holds in the triangle inequality only when x, y, z are colinear. In this case we would have a model for a geometry which satisfies all the axioms for hyperbolic geometry except the side-angle-side axiom. This geometry would nevertheless have a bona-fide metric satisfying the strict triangle inequality.

In defense of our apparently heartless abandonment of the side-angle-side axiom, we offer this quote of Choquet [7, p. 79] concerning the "unwarranted universal use" of angles. "There are strong reasons for doing away with this and developing almost the whole of our geometry without measuring a single angle. Although it is an essential tool of analysis and applied mathematics, angle is often

merely baggage, as far as geometry is concerned, and sometimes a source of error."

6. Some questions. The part metric is defined in terms of the positive harmonic functions in quite general domains. In particular, if G is a plane domain whose complement has at least two points then there are non-constant positive harmonic functions, and d is a conformally invariant metric. There are other conformally invariant metrics on general domains, such as the Bergman metric and the Carathéodory metric. Are they different?

For the disc, the gradients of the minimal functions fill up the domain nicely, so that for any two points z and w there is exactly one gradient line (one hyperbolic line) through z and w. The h-line (α, β), which is the gradient for both P_α and P_β, is furnished with a coordinate function $\log P_\alpha$, or equivalently $\log P_\beta$, which gives a hyperbolic distance. Does this happen in more general domains? Since minimal functions are preserved by conformal maps of any domain on itself, any such domain would provide a model for hyperbolic geometry. Are there multiply connected models for hyperbolic geometry?

In linear programming one maximizes a linear function on some convex set. This maximum will always occur at an extreme point, and the art consists of finding one's way as quickly as possible to the right extreme points. The convex sets in question will generally have flat boundaries, so the part metric will satisfy $d(x, y) + d(y, z) = d(x, z)$ not only when x, y, and z are colinear, but when the lines x, y and y, z intersect on the boundary on the same segment. Is this property, or other properties of the part metric relevant to the search for an appropriate extreme point?

REFERENCES

1. Heinz Bauer and H. S. Bear, The part metric in convex sets, *Pac. J. Math.*, 30 (1969), 15–33.
2. H. S. Bear, A geometric characterization of Gleason parts, *Proc. Am. Math. Soc.*, 16 (1965), 407–412.
3. H. S. Bear and M. L. Weiss, An intrinsic metric for parts, *Proc. Am. Math. Soc.*, 18 (1967), 812–817.
4. G. D. Birkhoff, A set of postulates for plane geometry based on scale and protractor, *Ann. of Math.*, 33 (1932), 239–245.
5. Herbert Busemann, The Geometry of Geodesics, Academic Press, New York, 1955.
6. C. Carathéodory, Conformal Representation, Cambridge Univ. Press, Cambridge, 1963.
7. Gustave Choquet, Geometry in a Modern Setting, Herman, Paris, 1969.
8. Saunders MacLane, Metric postulates for plane geometry, *Am. Math. Month.*, 66 (1959), 543–555.
9. R. S. Martin, Minimal positive harmonic functions, *Trans. Am. Math. Soc.*, 49 (1941), 137–172.
10. Edwin Moise, Elementary Geometry from an Advanced Standpoint, 2nd ed., Addison-Wesley, Reading, 1974, p. 113.
11. H. L. Royden, Real Analysis, 2nd Ed., Macmillan Co., 1968.
12. Walter Rudin, Real and Complex Analysis, 2nd Ed., McGraw Hill, New York, 1966.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Arcs with Positive Measure and a Space-Filling Curve

TIMOTHY LANCE AND EDWARD THOMAS
Department of Mathematics and Statistics, State University of New York, Albany, NY 12222

In this note we construct a one-parameter family of arcs in R^2 with positive two-dimensional Lebesgue measure, which converge uniformly to a space-filling curve. By an arc we mean a homeomorphism of $I = [0, 1]$ into $I \times I$. The existence of such maps is not news, but our procedure for constructing them is so transparently simple that it can be presented in undergraduate or graduate courses in advanced calculus, real analysis, or topology. Furthermore, each of these curves is the uniform limit of a recursively defined sequence of very simple functions. The examples of space filling curves of Peano or Hilbert, which are generally cited (e.g., [1], [2], or [3]), are also constructed recursively, but the recursions are more complicated. All of them pass from one stage to the next by splicing together reduced and rotated copies of an initial figure. In our construction, no rotation is necessary to get them to fit, so the recursion has a much simpler description. Our construction is similar to one found in Section 10.10 of [1]. This work was motivated by problem E 2975, posed by F. B. Jones in the December, 1982, issue of the *American Mathematical Monthly*. We wish to thank the referee for several helpful suggestions.

Let us first describe a typical member of our family of arcs intuitively. Imagine a necklace composed of four congruent squares and three connecting line segments arranged as in FIGURE 1. In each square we put another necklace, but one in which the "+" shaped region is proportionately much thinner. By leaving the original

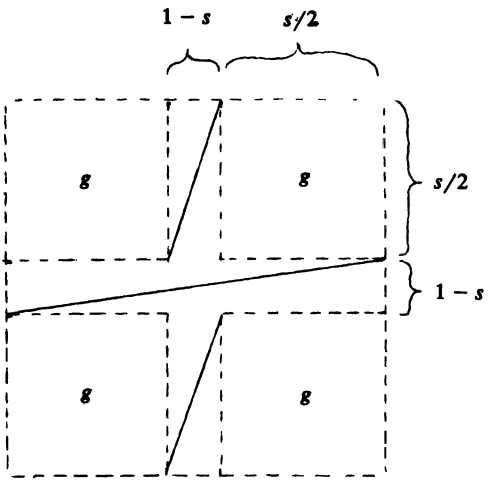


FIG. 1. The recursion operator R_s.

connecting line segments in place, we obtain a necklace with sixteen squares and fifteen segments. Iterating this process yields a family of necklaces whose intersection is an arc. (To prove this, note that the intersection A of the necklaces will be compact, connected, locally connected and, therefore, arcwise connected [2]. Let B be an arc in A which joins the northwest corner to the southeast corner. Clearly B contains all the connecting segments in each necklace. These are dense in A, hence $B = A$.)

The arc we have constructed contains the two-dimensional Cantor set determined by the system of squares. By choosing the squares appropriately the Cantor set—and hence the arc itself—can have any preassigned two-dimensional Lebesgue measure. By varying the system of squares continuously, we obtain a one parameter family of arcs whose limit will be a space-filling curve. Intuitively, the isotopy of these arcs to a space-filling curve is obtained by squeezing all the area out of the "+" shaped regions.

We begin by defining, for each $s \in (0, 1]$, the replication operator R_s. Suppose $g: I \to I \times I$ is any path with $g(0) = (0, 1)$ and $g(1) = (1, 0)$. Then $R_s(g): I \to I \times I$ is the path represented schematically by the following diagram.

Thus on each of the subintervals $[0, 1/7]$, $[2/7, 3/7]$, $[4/7, 5/7]$, and $[6/7, 1]$, $R_s(g)$ traces out a reduced and translated copy of the path g, and it is the indicated straight line paths on the remaining three subintervals. For any $s \in (0, 1]$ we define maps $f_{s,n}: I \to I \times I$ recursively by letting $f_{s,0}$ be the path $t \to (t, 1 - t)$ tracing one of the diagonals of $I \times I$ and setting

$$f_{s,n} = R_{(1+s)/2}(f_{2s/(1+s),n-1}).$$

For example, we have the following graphs for $s = 1/2$.

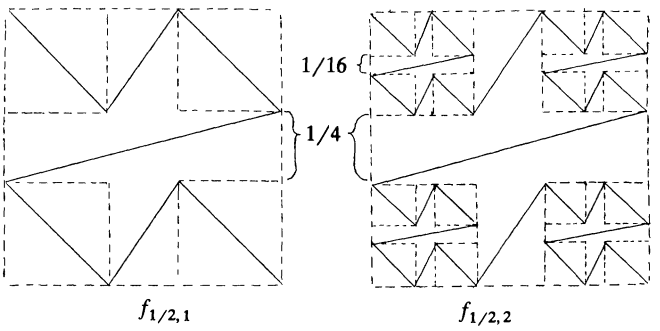


FIG. 2. The first two iterations for $f_{1/2}$.

THEOREM. *For each* $s \in (0, 1]$ *and* $t \in [0, 1]$ *the limit*

$$f_s(t) = \lim_{n \to \infty} f_{s,n}(t)$$

exists and defines a continuous map $(0, 1] \times I \to I \times I$. *If* $0 < s < 1$ *the path* $t \to f_s(t)$ *is an arc with Lebesgue measure* s^2. *The path* $t \to f_1(t)$ *is a space-filling curve in* $I \times I$ *which is the uniform limit of the arcs* $t \to f_s(t)$.

Proof. The maps $f_{s,n}$ and $f_{s,m}$, $n < m$, agree except on 4^n subintervals of I each of length $(1/7)^n$. On each of those subintervals their values lie in a square at

most $(1/2)^n$ on a side, and the center of that square clearly varies continuously in s. The existence and continuity of $(s, t) \to f_s(t)$ follows. If $s < 1$ each of the maps $t \to f_{s,n}(t)$ is one to one. Thus if $|t_1 - t_2| > 1/7^n$, then either t_1 and t_2 both lie in an interval where $f_{s,n} = f_s$ or they are sent by $f_{s,n}$ (and hence by f_s) to different regions of $I \times I$. Thus $t \to f_s(t)$ is an arc for $0 < s < 1$.

Each application of operator R divides a given square into four smaller squares and a cross-shaped region in between, and in that region the graphs of $f_{s,n}$ and, hence, the graph of f_s clearly have measure 0. Thus if A_n denotes the union of the 4^n squares in $I \times I$ used in defining $f_{s,n}$, then the Lebesgue measure of $f_s(I)$ equals that of $\bigcap_{n=1}^{\infty} A_n$. But it follows inductively that A_n has measure $((1 + (2^n - 1)s)/2^n)^2$. When $s = 1$ the curve $t \to f_1(t)$ not only has measure 1 but its image contains each of the lines $x = k/2^n$ and $y = k/2^n$ for $k = 1, 2, \ldots, 2^n - 1$ and $n = 1, 2, 3, \ldots$. Thus $f_1(I) = I \times I$. □

Although a variety of different recursions will yield arcs of positive measure, one must be sure that the cross-shaped regions are squeezed to zero fast enough. For example, even if g_0 is a space-filling curve, the recursion $g_n = R_s(g_{n-1})$ converges to an arc of measure 0 whenever $s < 1$. When $s = 1$, however, we do get the simple recursion

$$f_{1,n} = R_1(f_{1,n-1}).$$

The resulting sequence of functions which converges to the space filling curve f_1 begins as follows. Notice that these curves are partially collapsed versions of the first three iterations in FIGURE 2.

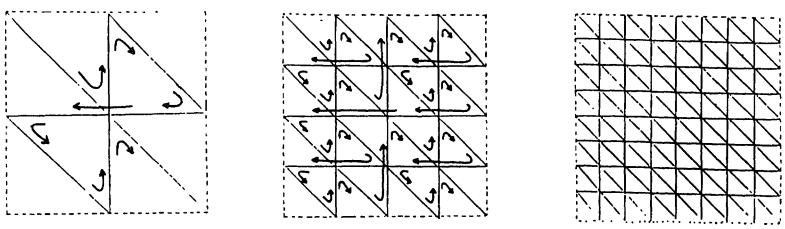


FIG. 3. Three iterations for the space filling curve f_1.

Finally, we note that all of the maps above are related by continuous deformations of the entire square $I \times I$. Fix some $s \in (0, 1]$, and consider the Cantor set C_s obtained by removing an open interval of length $(1 - s)/2$ from the middle of I, open intervals of length $(1 - s)/8$ from the center of the two resulting open intervals, open intervals of length $(1 - s)/32$ from the four resulting intervals, and so on. The Cantor set C_s has one dimensional Lebesgue measure s. There is a standard mapping of C_s onto I, the Cantor function Φ_s (see [1] or [2]), which can be embedded as the last stage of a one parameter family of maps $t \to \Phi_{s,t}$. In this family, $\Phi_{s,0}$ is the identity, $\Phi_{s,t}$ is a homeomorphism for $t < 1$, and $\Phi_{s,1}$ restricted to C_s equals Φ_s. It is possible to write formulas for $\Phi_{s,t}$, but it is simpler to draw a picture which illustrates the dynamics.

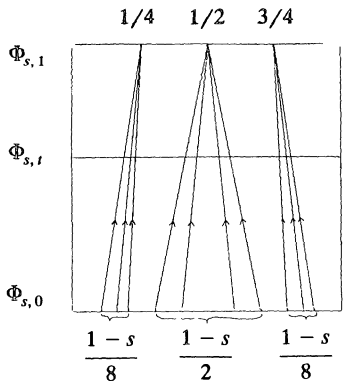


FIG. 4. A homotopy from the identity to the Cantor function.

The triangular regions are chosen with their top vertices at rational numbers of the form $(2k + 1)/2^m$. Thus if we delete the interior of all such triangles, the intersection of the resulting set with the line $y = t$ is precisely the Cantor set $C_{t+s(1-t)}$ described above. Define a two-parameter family of maps $F_{s,t}: I \times I \to I \times I$ by $F_{s,t}(x, y) = (\Phi_{s,t}(x), \Phi_{s,t}(y))$ for $s \in (0, 1]$ and $t, x, y \in [0, 1]$. This is clearly jointly continuous in all four variables. For fixed s the family $F_{s,t} \circ f_s$ is precisely the family $f_{t+s(1-t)}$. Running this homotopy backwards gives us an unfolding of $I \times I$ into the arc f_s with Lebesgue measure s^2.

REFERENCES

1. B. Gelbaum and J. Olmstead, Counterexamples In Analysis, Holden-Day, Inc., San Francisco, 1964.
2. J. Hocking and G. Young, Topology, Addison-Wesley Publishing Co., Reading, Mass., 1961.
3. J. Munkres, Topology, A First Course, Prentice Hall, Inc., Englewood Cliffs, N.J., 1975.

Unions and Common Complements of Subspaces

K. P. S. BHASKARA RAO
Department of Mathematics, Wesleyan University, Middleton, CT 06457

A. RAMACHANDRA RAO
Indian Statistical Institute, Calcutta-35, India

The following exercise can be found in most undergraduate text-books on linear algebra: show that if S and T are subspaces of a vector space, $S \cup T$ is a subspace if and only if either $S \subseteq T$ or $T \subseteq S$. In our experience, most of the students are able to show this and a few get curious about what happens for three or more subspaces.

The most natural conjecture for the general case would be that the union of a family $\mathcal{S}$ of subspaces of a vector space V over a field F is itself a subspace of V

if and only if some member of $\mathscr{S}$ contains all the others. But one can immediately see that this is false since any vector space with dimension at least two is the union of its subspaces with codimension 1 and none of these can contain the others. So we raise the question: can we impose some reasonable restrictions on the cardinalities of $\mathscr{S}$ and F and the dimensions of the subspaces in $\mathscr{S}$ to make the conjecture true? The answer is, indeed, yes. For example, the conjecture is known to be true if $\mathscr{S}$ is assumed to be finite and F infinite, see Problem 21, p. 177 of [**2**] or Theorem 1 below. In this note we present other reasonable conditions under which the conjecture is true. We also study the existence of a common complement for a family of subspaces, which turns out to be related to the above problem, and deduce a result of Lord [**4**].

To set up the notation let V be a (not necessarily finite dimensional) vector space over a (not necessarily infinite) field F. We will use $|A|$ to denote the cardinality of set A. For a subspace S, $\dim(S)$ denotes the dimension of S while $\operatorname{cod}(S)$ denotes the codimension of S. It may be noted that $\operatorname{cod}(S)$ is the dimension of any complement of S and equals $\dim(V/S)$. $\operatorname{Sp}(X)$ will denote the linear span of the set X of vectors.

To start with we will show that for any vector space V over F with dimension at least 2, we must have $|I| < |F| + 1$ for the conjecture to have any chance of holding. To see this, fix a basis $\{x_i\colon i \in I\}$ of V and two distinct elements j and k of I. Consider the $|F| + 1$ subspaces $W_\alpha = \operatorname{Sp}(\{x_j + \alpha x_k\} \cup \{x_i\colon i \neq j, k\})$ for all $\alpha \in F$ and $W_\infty = \operatorname{Sp}(\{x_i\colon i \neq j\})$. It is easy to check that these are distinct subspaces with codimension 1 (so none of them can contain the others) and that their union is V.

However, $|I| < |F| + 1$ alone is not sufficient: consider the vector space V of all polynomials over $\mathbb{R}$ and the subspaces $S_0, S_1, S_2, \ldots$ where S_i consists of all polynomials with degree at most i—clearly no S_i contains all the others though their union is V. Thus, to prove the conjecture, we have to make some further assumption: for example, the family is finite or the subspaces are of bounded dimension. We now proceed to prove the conjecture under each of these assumptions.

THEOREM 1. *Let $S_1, S_2, \ldots, S_k$ be finitely many subspaces of V with $k < |F| + 1$. Then $S_1 \cup S_2 \cup \cdots \cup S_k$ is a subspace if and only if some S_i contains the others.*

Proof. The "if" part of the theorem is obvious. We prove the other part by induction on k. The case $k = 1$ is trivial. We now assume the result for $k - 1$ and prove it for k. So let $S_1 \cup \cdots \cup S_k$ be a subspace and $k < |F| + 1$. If $S_1 \supseteq S_2 \cup \cdots \cup S_k$, we are through. So let $S_1 \not\supseteq S_2 \cup \cdots \cup S_k$. We shall show that $S_1 \subseteq S_2 \cup \cdots \cup S_k$. To this end, fix a $y \in (S_2 \cup \cdots \cup S_k) - S_1$ and let $x \in S_1$. Then for every scalar α, $\alpha x + y \notin S_1$. Since $S_1 \cup \cdots \cup S_k$ is a subspace, $\alpha x + y \in S_2 \cup \cdots \cup S_k$. Since $\alpha x + y \neq \beta x + y$ whenever $\alpha \neq \beta$ and $|F| \geqslant k$, it follows that there exist j with $2 \leqslant j \leqslant k$ and $\alpha \neq \beta$ in F such that $\alpha x + y$ and $\beta x + y$ are in S_j. Then $x \in S_j$. Thus $S_1 \subseteq S_2 \cup \cdots \cup S_k$. Now the result follows from the induction hypothesis. ■

We now consider the case of a finite-dimensional vector space before going on to subspaces with bounded dimension in an infinite-dimensional vector space. We state only the non-trivial part in the remaining theorems.

THEOREM 2. *Let* $\{S_i: i \in I\}$ *be a family of subspaces of a finite-dimensional vector space* V *with* $|I| < |F| + 1$. *If* $S := \bigcup_{i \in I} S_i$ *is a subspace then* $S = S_i$ *for some* $i \in I$.

Proof. We prove the theorem by induction on n, the dimension of S. For $n = 1$ the result is easy to prove. Next assume the result for $n - 1$ and let $n \geqslant 2$.

Suppose that no S_i equals S. If H is *any* subspace of S with dimension $n - 1$,

$$H = H \cap S = H \cap \left(\bigcup_{i \in I} S_i \right) = \bigcup_{i \in I} (H \cap S_i),$$

so by the induction hypothesis, $H = H \cap S_j$ for some $j \in I$. Then $H \subseteq S_j$ and S_j is a proper subspace of S. Since $\dim(H) = \dim(S) - 1$ it follows that $H = S_j$. Thus *every* subspace of S with dimension $n - 1$ is one of the S_i's. Since, as we saw in the first example before Theorem 1, there are at least $|F| + 1$ such subspaces of S, we get $|I| \geqslant |F| + 1$, a contradiction which proves the theorem. ∎

For a denumerable family Theorem 2 was proved earlier by Byrd [**1**].

We now deduce the result for a family of subspaces with bounded dimension.

THEOREM 3. *Let* $\{S_i: i \in I\}$ *be a family of subspaces of* V *with* $|I| < |F| + 1$. *Also let* m *be a positive integer such that* $\dim(S_i) \leqslant m$ *for all* $i \in I$. *If* $S := \bigcup_{i \in I} S_i$ *is a subspace then* $S = S_i$ *for some* $i \in I$.

Proof. We first show that $\dim(S) \leqslant m$. Suppose not. Then there exists a subspace W of S with dimension $m + 1$. Applying Theorem 2 to the family $\{W \cap S_i: i \in I\}$ of subspaces of W, we get $W = W \cap S_i$ for some $i \in I$, a contradiction since $\dim(W) = m + 1$ and $\dim(S_i) \leqslant m$. Now an application of Theorem 2 to the family $\{S_i: i \in I\}$ of subspaces of S yields the present theorem. ∎

We now use the above results to give conditions under which a family of subspaces of a vector space has a common complement.

Observe that if the S_i's have a common complement then they must have the same dimension and the same codimension. We restrict our attention to families of subspaces with common finite dimension or common finite codimension and we start with the latter case which happens to be easier.

THEOREM 4. *Let* $\{S_i: i \in I\}$ *be a family of subspaces of* V *with* $|I| < |F| + 1$. *If the* S_i*'s have a common finite codimension* p *and either* I *is finite or* V *is finite-dimensional, then the* S_i*'s have a common complement.*

Proof. We prove the theorem by induction on p. If $p = 0$, $\{0\}$ is a common complement of the S_i's. Next assume the result for codimension $p - 1$ and let $\operatorname{cod}(S_i) = p$ for all $i \in I$, where $p \geqslant 1$. Then by Theorems 1 and 2, $\bigcup_i S_i \neq V$. Fix an x in $V - \bigcup_i S_i$ and consider $T_i = \operatorname{Sp}(S_i \cup \{x\})$ for all $i \in I$. Clearly $\operatorname{cod}(T_i) = p - 1$ for all i, so by the induction hypothesis, the T_i's have a common complement W. $\operatorname{Sp}(W \cup \{x\})$ is then a common complement of the S_i's. ∎

COROLLARY 5 (Lord [**4**]). *Let* V *be an* n*-dimensional vector space over* $\mathbb{R}$ *or* $\mathbb{C}$ *and let* $\{S_i: i \in \mathbb{N}\}$ *be a countable family of subspaces of* V *each having the same dimension. Then the* S_i*'s have a common complement.*

The proof of Corollary 5 essentially appears in Byrd [**1**].

THEOREM 6. *Let $\{S_i: i \in I\}$ be a family of subspaces of V with $|I| < |F| + 1$. If the S_i's have a common finite dimension, they have a common complement.*

Proof. If V is finite-dimensional, this follows from Theorem 4. Otherwise, applying Zorn's lemma to the family $\mathcal{T}$ of all subspaces T of V with the property $T \cap S_i = \{0\}$ for all i gives a maximal element T_0 of $\mathcal{T}$. We now prove that T_0 is a common complement of the S_i's. Suppose T_0 is not a complement of S_j. Let φ be the natural homomorphism from V to V/T_0. Since $T \cap S_i = \{0\}$ it follows that $\dim(\varphi(S_i)) = \dim(S_i) = m$ (say) for all $i \in I$. Since T_0 is not a complement of S_j it follows that $\varphi(S_j)$ is a proper subspace of V/T_0, so $\dim(V/T_0) > m$. So by Theorem 3, $\bigcup_i \varphi(S_i) \neq V/T_0$. Hence, there is a 1-dimensional subspace X of V/T_0 such that $X \cap \varphi(S_i) = \{0\}$ for all i. Now $\varphi^{-1}(X)$ is a member of $\mathcal{T}$ properly containing T_0, a contradiction which proves that T_0 is a common complement of the S_i's. ∎

We now show that when $\dim(V) \geqslant 2$, the condition $|I| < |F| + 1$ cannot be dropped in Theorems 4 and 6. Regarding Theorem 4 it is enough to consider (again) $|F| + 1$ subspaces of V with codimension 1 and union V. Regarding Theorem 6 we first choose a subspace Y of V with dimension 2. Then we consider a family $\{S_i: i \in I\}$ of $|F| + 1$ subspaces of Y with common dimension 1 and union Y. Suppose the S_i's have a common complement T in V. Then $S_i + (T \cap Y) = (S_i + T) \cap Y = Y$ for all i, so $T \cap Y$ is a common complement of the S_i's in Y, an impossibility which proves that the S_i's cannot have a common complement in V.

It is easy to see that the condition "either I is finite or V is finite dimensional" in Theorem 4 cannot be dropped. For this, take a basis $B = \{x_j: j \in J\}$ of an infinite-dimensional vector space V and consider $\{\mathrm{Sp}(B - \{x_i\}): i \in I\}$ for any countably infinite subset I of J. Indeed, the same example also shows that "dimension" in Theorem 6 cannot be replaced by "codimension."

We end this note by answering the problems raised in [3] and [4]. In [3] the problem was raised as to what are the results analogous to Corollary 5 which hold for vector spaces over finite fields. The answer is provided by our Theorems 4 and 6.

The problem raised in [4] was what is the maximum number of distinct m-dimensional subspaces of an n-dimensional vector space V over a finite field F that have a common complement? The answer is the number of distinct complements of an $(n - m)$-dimensional subspace T since this number is the same for all T. To find this number, fix an ordered basis B of T. Any complement of T can be obtained as $\mathrm{Sp}(C)$ where $B \cup C$ is an extension of B to a basis of V. This extension can be done in

$$(|F|^n - |F|^{n-m})(|F|^n - |F|^{n-m+1}) \cdots (|F|^n - |F|^{n-1})$$

ways. But different C's can give rise to the same complement of T. The number of C's giving rise to the same complement is the same as the number of bases of an m-dimensional subspace and this is $(|F|^m - 1)(|F|^m - |F|) \cdots (|F|^m - |F|^{m-1})$. Thus the required number is

$$\frac{(|F|^n - |F|^{n-m})(|F|^n - |F|^{n-m+1}) \cdots (|F|^n - |F|^{n-1})}{(|F|^m - 1)(|F|^m - |F|) \cdots (|F|^m - |F|^{m-1})} = |F|^{m(n-m)}.$$

Acknowledgement. We thank one of the referees for some suggestions regarding the presentation of the material and for bringing [1] to our notice.

REFERENCES

1. R. D. Byrd, Simultaneous complements in finite-dimensional vector spaces, *Amer. Math. Monthly*, 93 (1986) 641–642.
2. I. N. Herstein, Topics in Algebra, second edition, 1975.
3. B. Levinger, Review of [4] in *Math. Reviews*, 87b: 15003.
4. N. J. Lord, Simultaneous complements in finite-dimensional vector spaces, *Amer. Math. Monthly*, 92 (1985) 492–493.

The Generation of All Rational Orthogonal Matrices

HANS LIEBECK AND ANTHONY OSBORNE
Department of Mathematics, University of Keele, Staffordshire, ST5 5BG, England

In a letter to this MONTHLY [1] John Cremona shows how to generate all 3×3 orthogonal matrices with rational coefficients. His method is based on the real algebra of quaternions. By a further application of quaternions one can obtain 4×4 rational orthogonal matrices. (See du Val [2].)

In this note we show how to generate all $n \times n$ rational orthogonal matrices and hence all orthonormal bases of the rational vector space $\mathbf{Q}^n$. At the same time we obtain all real orthogonal matrices and all complex unitary matrices. Our method is based on a further piece of mathematics from the last century—Cayley's formula for orthogonal matrices.

Preliminaries. Let A and B be $n \times n$ matrices. We shall say that B *is equivalent to* A, and write $B \sim A$, if and only if there exists a diagonal matrix D with diagonal entries selected from the set $\{-1, 1\}$ such that $B = DA$. Clearly B is equivalent to A if and only if, for each $i = 1, \ldots, n$, the ith row of B is $\pm i$th row of A. The relation $\sim$ is an equivalence relation on the set of $n \times n$ matrices, and the equivalence class $\mathscr{E}(A)$ which contains A has at most 2^n members.

LEMMA. *Let A be an $n \times n$ matrix over a field of characteristic $\neq 2$. Then at least one of the matrices in $\mathscr{E}(A)$ does not have eigenvalue* 1.

Proof. By induction on n. For $n = 1$ the result is true, since the 1×1 matrices A and $-A$ cannot both be [1] over a field of characteristic $\neq 2$. Proceeding by way of contradiction let n be the least order for which the lemma is false. Then there exists an $n \times n$ matrix A such that all matrices in $\mathscr{E}(A)$ have eigenvalue 1. Thus for any matrix $DA \in \mathscr{E}(A)$, there exists $\mathbf{x} \neq \mathbf{0}$ such that $DA\mathbf{x} = \mathbf{x}$. Hence $A\mathbf{x} = D\mathbf{x}$, and so the matrix $A - D$ is singular. Consider the determinant function

$$d(t_1, \ldots, t_n) = \det(A - \operatorname{diag}(t_1, \ldots, t_n)).$$

We have that $d(t_1, \ldots, t_n) = 0$ for all 2^n choices of $t_i = \pm 1$, $i = 1, \ldots, n$.

Expanding $d(t_1, \ldots, t_n)$ according to the first row, we obtain

$$d(t_1, \ldots, t_n) = (a_{11} - t_1)d^*(t_2, \ldots, t_n) + \text{terms independent of } t_1. \quad (1)$$

By the induction hypothesis the determinant $d^*(t_2, \ldots, t_n)$ of order $n-1$ is non-zero for at least one choice of $t_j = \pm 1$, $j = 2, \ldots, n$. Make such a choice: $t_2, \ldots, t_n = \lambda_2, \ldots, \lambda_n$.

We have $d(1, \lambda_2, \ldots, \lambda_n) = 0$ and $d(-1, \lambda_2, \ldots, \lambda_n) = 0$. From (1)

$$(a_{11} - 1)d^*(\lambda_2, \ldots, \lambda_n) = (a_{11} + 1)d^*(\lambda_2, \ldots, \lambda_n).$$

Hence $a_{11} - 1 = a_{11} + 1$ which is clearly false in a field of characteristic $\neq 2$. So the lemma is proved.

Construction of $n \times n$ orthogonal matrices. Let F be a subfield of the field of real numbers. By Cayley's formula [3 p. 289] the mapping $S \mapsto (S - I)^{-1}(S + I)$ gives a one-one correspondence between the set of $n \times n$ skew-symmetric matrices S over F and the set of $n \times n$ orthogonal matrices over F which do not have 1 as an eigenvalue. The inverse correspondence is given by $A \mapsto (A - I)^{-1}(A + I)$.

Now let A be any $n \times n$ orthogonal matrix over F. Then the equivalence class $\mathscr{C}(A)$ consists of orthogonal matrices, and by the Lemma at least one of these, B say, does not have eigenvalue 1. By the Cayley correspondence there exists a skew-symmetric matrix S over F such that $B = (S - I)^{-1}(S + I)$. Thus we have the following result.

THEOREM. *The set of all $n \times n$ orthogonal matrices over F is generated by the $n \times n$ orthogonal matrices $(S - I)^{-1}(S + I)$, where S is skew-symmetric over F, and their equivalence classes.*

The same technique can be used to generate all unitary matrices by considering the complex field and replacing "orthogonal" by "unitary" and "skew-symmetric" by "skew-hermitian" in the above analysis.

It is worth pointing out that the $n \times n$ orthogonal matrices $(S - I)^{-1}(S + I)$ referred to in the theorem all have determinant $(-1)^n$. (Proof: $\det(S + I) = \det(S + I)^T = \det(-S + I) = (-1)^n \det(S - I)$, from which the result follows.) So they represent reflections when n is odd and rotations when n is even. Consequently every $n \times n$ reflection matrix when n is even and every $n \times n$ rotation matrix when n is odd is equivalent but not equal to a matrix of the form $(S - I)^{-1}(S + I)$. This can also be seen from [4] Theorem B, which states that every reflection of a (Euclidean) space of even dimension and every rotation of a space of odd dimension has a fixed vector. The corresponding orthogonal matrices then have eigenvalue 1 and, therefore, do not feature in the Cayley correspondence.

In particular, John Cremona's 3×3 rotation matrix $M(a, b, c, d)$ is given for the case $d \neq 0$ by

$$M(a, b, c, d) = D(S - I)^{-1}(S + I),$$

where S is the 3×3 skew-symmetric matrix which has a/d, $-b/d$, $-c/d$ above the diagonal, and $D = \operatorname{diag}(1, 1, -1)$.

This note provides a theoretical method of generating all orthogonal matrices over a real field F, but as a practical classroom aid it is of limited use. The reader interested in a simple technique of generating a large class of 3×3 rational orthogonal matrices is referred to our Teaching Note [5].

We are grateful to a referee for some of the observations following our theorem.

REFERENCES

1. John Cremona, Letter to the Editor, *Amer. Math. Monthly*, 94 (1987) 757.
2. Patrick du Val, Homographies, Quaternions and Rotations, Oxford University Press, 1964.
3. F. R. Gantmacher, The Theory of Matrices, Vol. I, Chelsea, 1960.
4. A. M. Adelberg, Reflections have reversed vectors, *Amer. Math. Monthly*, 79 (1972) 59–62.
5. Anthony Osborne and Hans Liebeck, Orthogonal bases of $\mathbb{R}^3$ with integer coordinates and integer lengths, *Amer. Math. Monthly*, 96 (1989) 49–53.

Hypocycloids, Continued Fractions, and Distribution Modulo One

NORMAN RICHERT*
Division of Mathematics and Systems Design, University of Houston, Clear Lake, Houston, TX 77058

In simulation and Monte Carlo approximation schemes, a random sequence (pseudorandom, really) is needed. How this sequence might be generated is a long story. Donald Knuth [2] encapsulates some of the difficulties neatly in *The Art of Computer Programming, Volume 2*, where he says that "*random numbers should not be generated with a method chosen at random*. Some theory should be used." A sequence with some very nice properties is the so-called $(n\theta)$ sequence. Let θ be irrational. Let a sequence $\{x_n\}$ in $[0, 1)$ be defined by $x_n = (n\theta)$, $n = 1, 2, \ldots$, where $(\cdot)$ denotes the fractional part. Let us explore this sequence.

One way to visualize what the sequence is doing is to imagine a circle of circumference θ rolling to the right on the real line. The circle has a distinguished point P, which begins at zero. Then the sequence consists of the set of positive points where P touches the line, as long as each interval from integer k to $k + 1$ is understood as a copy of $[0, 1)$. So the points could be viewed as the set of cusp points of the cycloid. Better still, roll the real line up into a circle of circumference 1 by identifying the integer points. Now the cycloid is a hypocycloid, at least if the circumference of the rolling circle is less than 1. There is no loss of generality in supposing this, so assume in what follows that $0 < \theta < 1$.

Calculus students are frequently asked to find parametric equations for a hypocycloid, using the angle between the x-axis and the line of centers of the two circles as the parameter. Frequently, $\theta = .25$ is illustrated as a special case, the hypocycloid of four cusps, because the parametric equations have a particularly simple form in that case. It is easy to infer a family of hypocycloids of q-cusps, where $\theta = 1/q$, $q \geqslant 2$. The associated sequences could hardly be described as random. From a theoretical point of view, things get more interesting if θ is irrational. Then the curve is not closed and $\{(n\theta)\}$ is not a periodic sequence. However, on a computer all numbers are rational, and plotters have only a finite resolution. So for the purposes of this discussion the rational/irrational distinction

*Appreciation is expressed to Marquette University Computer Services for the use of the large-format drum-pen plotter used in the figures.

will be blurred. We might view irrational numbers as rationals with large, unknown denominators.

If $\theta = p/q$ is rational, in reduced form, it is not hard to see that the rolling circle must roll completely around p times to bring P back to the 0 point, with q cusps. If $p = 1$, a simple closed curve results (except in the degenerate case of $q = 2$); otherwise the curve is closed, but not simple. So the hypocycloids of q-cusps form a somewhat larger family than it seems at first glance. In particular, there are $\phi(q)$ such curves, where ϕ denotes the Euler ϕ-function. However, if we consider the hypocycloid only as a point set, the *trace* of the hypocycloid, we have over counted by a factor of 2. As a comparison of the curves associated with $\theta = .25$ and $\theta = .75$ suggests, the traces associated with θ and $1 - \theta$ are the same.

Let us introduce a bit of notation. Consider the parameterized curves as functions γ_θ: $(0, \infty) \to \mathbf{R}^2$. Denote the trace of γ_θ by $\{\gamma_\theta\}$. The domain $(0, 2p\pi]$ is convenient for plotting $\{\gamma_\theta\}$ with no overlap where $\theta = p/q$. Let $\{\gamma_\theta\}^t, t \in (0, q]$, denote the trace of γ restricted to $(0, 2\pi tp/q]$, so that $\{\gamma_\theta\}^t$ can be thought of as a "partially completed" plot of γ_θ. Let $(n\theta)^t$ be the finite sequence associated with $\{\gamma_\theta\}^t$, that is, $\{(n\theta)\}_{n=1}^{[t]}$, where $[\cdot]$ denotes the greatest integer function. For $\theta = 1/q$, the various $\{\gamma_\theta\}^t$ are easily predictable, but for $p > 1$ the situation is less clear. Various patterns of cusps seem to appear as t increases. Moreover, these patterns do not seem to be continuously related to θ. Compare FIGURE 1 with $\theta = 101/251$ and $t = 169$ with FIGURE 2, where $\theta = 102/251$ and $t = 155$. The

$\theta = 102/251$

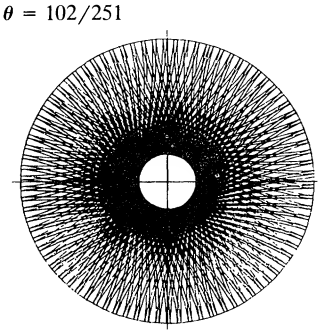


FIG. 1.

$\theta = 101/251$

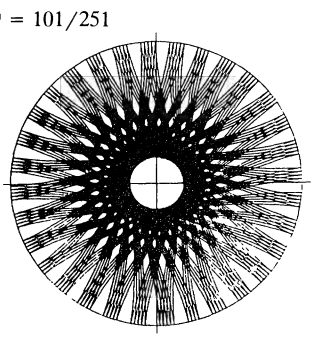


FIG. 2.

cusps in the latter case group into 32 clumps, a phenomenon absent in the first case (additional figures with other t values in each case would be more convincing). "Clump" is not a technical term, but I believe the figures show what is meant. Without going into the details, it seems reasonable that $(n(102/251))^{155}$ will fare worse in statistical tests that emphasize uniformity of distribution than will $(n(101/251))^{169}$, worse than the differing number of points will account for.

The clumping noted in the figures occurs when the tracing point P comes closer to 0 than it previously has, i.e., when $\min\{(n\theta), 1 - (n\theta)\}$ attains a minimum. The trace from that point onwards is a copy of the initial trace, with a small rotation. The smaller this minimum is, the more the clumping. So we need to investigate the minimal values of $|n\theta - m|$ for integral m, n. Continued fractions [5] are exactly the mechanism to investigate these minimal values. The values $|q_k\theta - p_k|$ corresponding to the sequence of *convergents* p_k/q_k associated with θ are exactly these minimal values.

The continued fraction expansion of θ consists of a sequence of integers a_k, positive for $k > 0$, the *partial quotients* (finite when θ is rational) and an associated sequence of numbers ξ_k, the *complete quotients* defined by

$$\xi_0 = \theta, \qquad a_0 = [\xi_0],$$

$$\xi_k = \frac{1}{\xi_{k-1} - a_{k-1}} \qquad a_k = [\xi_k], \qquad k \geqslant 1.$$

The partial quotients yield the convergents, given by

$$p_{-2} = 0, \quad p_{-1} = 1, \quad p_k = a_k p_{k-1} + p_{k-2},$$

$$q_{-2} = 1, \quad q_{-1} = 0, \quad q_k = a_k q_{k-1} + q_{k-2}, \quad k \geqslant 0.$$

One of the fundamental results in the theory of continued fractions [5] is that the convergents to a number θ are the *best approximations* to that number in the sense of minimizing $|b\theta - a|$ over all denominators no larger than b. The $22/7$ approximation of π is a convergent. FIGURE 3 shows $\{\gamma_\theta\}^{22}$ for $\theta = 1/\pi$, illustrating the $7/22$ convergent (if $\theta > 1$, the sequence of partial quotients for $\theta - [\theta] \to 1/\theta$ is simply shifted one place to the right, with $a_0 = 0$, and the convergents are

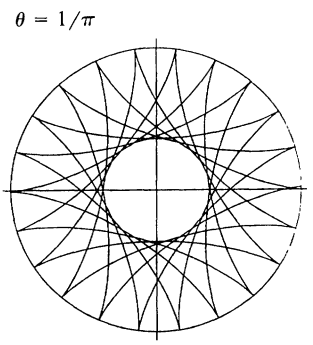


FIG. 3.

$\theta = 1/\pi$

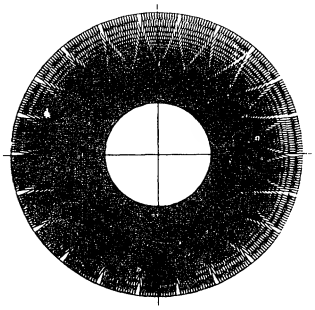


FIG. 4.

the reciprocals of the convergents for θ). FIGURE 4 is $\{\gamma_\theta\}^{333}$, illustrating the less familiar approximation, 106/333.

It is shown in the theory of continued fractions that the sequence $|q_k\theta - p_k|$ is decreasing, converging to 0. Thus the convergents p_k/q_k of θ correspond to nearly equally spaced sets of q_k cusps. When $|q_k\theta - p_k|$ is much smaller than the $1/q_k$ spacing between cusps, so that $q_k|q_k\theta - p_k|$ is small, there will be clumping. It can be shown that

$$q_k|q_k\theta - p_k| < \frac{q_k}{q_{k+1}} \leqslant \frac{1}{a_{k+1}},$$

which says for the hypocycloid pictures that a high level of clumping is associated with large partial quotients.

The partial quotients of the θ's in FIGURES 1 and 2 are $101/251 = [0; 2, 2, 16, 3]$ and $102/251 = [0; 2, 2, 5, 1, 7]$. In the latter case, the relatively large $a_5 = 7$ produces clumping around the cusps associated with $p_4/q_4 = 13/32$, and thus the 32 clumps in FIGURE 2. The partial quotient $a_3 = 16$ in the former case will yield clumping corresponding to $p_2/q_2 = 2/5$. This would be seen in $\{\gamma\}^t$ plots for t closer to 5.

The study of *uniformly distributed sequences modulo one* provides useful tools for investigating the clumping that shows up in these hypocycloid plots. Let E be a subset of $[0, 1)$ and for the sequence $\omega = \{x_n\}$ let $A(E; N; \omega) = \#\{n | 1 \leqslant n \leqslant N, (x_n) \in E\}$. Then ω is said to be *uniformly distributed modulo* 1 (u.d. mod 1) if for every pair of real numbers a, b with $0 \leqslant a < b \leqslant 1$ the equation

$$\lim_{N\to\infty} \frac{A([a, b); N; \omega)}{N} = b - a$$

holds. That is, each interval $[a, b)$ contains the appropriate portion of terms in the limit. It can be shown [4] that for θ irrational $(n\theta)$ is u.d. mod 1. The clumping in FIGURE 5 suggests that the sequence is not u.d. mod 1, but the limiting situation is all that matters in the definition. The tool used to measure finite segments of sequences is the *discrepancy*, D_N, defined by

$$D_N(\omega) = \sup_{0\leqslant\alpha<\beta\leqslant 1} \left| \frac{A([\alpha, \beta); N; \omega)}{N} - (\beta - \alpha) \right|.$$

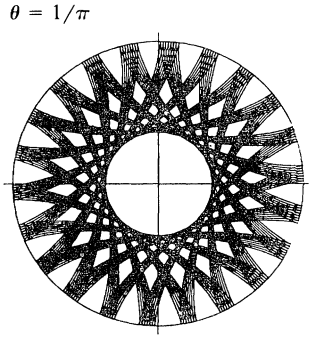


FIG. 5.

Clearly $D_N \leqslant 1$ and it can be shown that $D_N \geqslant 1/N$, with $D_N = 1/N$ if $\{x_n\}_{n=1}^N = \{(n-1)/N\}_{n=1}^N$. Visually, it seems clear that the sequence represented in FIGURE 2 has a higher discrepancy than the sequence of FIGURE 1. In fact, $D_{169} \doteq .010$ for $\theta = 101/251$ and $D_{155} \doteq .018$ for $\theta = 102/251$, so the former is closer to optimal, $1/169 \doteq .006$.

An easier measure to calculate is the *star discrepancy*, D_N^*, where the α in the definition of D_N is fixed at 0. It is shown in [6] that if $(x_1), \ldots, (x_N)$ are ordered by increasing magnitude, then D_N^* can be calculated by

$$D_N^* = \max_{i=1,\ldots,N} \max\left\{\left|\frac{i}{N} - (x_i)\right|, \left|\frac{(i-1)}{N} - (x_i)\right|\right\}.$$

This value satisfies $D_N^* \leqslant D_N \leqslant 2D_N^*$, so $1/(2N) \leqslant D_N^* \leqslant 1$, and $D_N^* = 1/(2N)$ iff $x_n = (2n-1)/2N$.

An important application of pseudorandom sequences is in Monte Carlo approximation methods. It is a result of Koksma [3] that if f is a function of bounded variation $V(f)$ on $[0, 1]$, then

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(t)\,dt\right| \leqslant V(f)D_N^*.$$

This application is an important source for interest in low discrepancy sequences. But why not simply always use $x_n = (2n-1)/2N$? First, this gives only a single sequence for each N, and a typical strategy in applying Koksma's approximation is to use several sequences, and perhaps average the results. Second, it may be desirable to add new points to an existing approximation to sharpen the results. But these must be added in multiples of N, since intermediate sequences would have relatively large discrepancies.

What is desired is a family of sequences for which D_N is as close as possible to $1/N$. It turns out that if θ is of a certain class of numbers, then $(n\theta)$ is just such a sequence. It is shown in [4, 6] that if θ has bounded partial quotients, then for every $\varepsilon > 0$, $D_N(\omega) = O(N^{-1}\log N)$. In fact, if θ has partial quotients bounded by K, then

$$ND_N(\omega) \leqslant 3 + \left(\frac{1}{\log \xi} + \frac{K}{\log(K+1)}\right)\log N,$$

where $\xi = (1 + \sqrt{5})/2$, and this is the least order of magnitude possible. So a particularly good sequence is provided by $\theta = \xi - 1 = (-1 + \sqrt{5})/2$, for which the continued fraction expansion after $a_0 = 0$ consists entirely of 1's.

What does this mean for the hypocycloid plots? For these "good" θ, the $\{\gamma_\theta\}'$ plots will never show the clumping seen in FIGURE 2, and also seen in FIGURE 5 with $\theta = 1/\pi$ and $t = 163.4$. FIGURES 6, 7 and 8 show $\{\gamma_{\xi-1}\}'$ for $t = 14.6$, 16.2, and 17.8. In a rough sense, the hypocycloids and their associated sequences divide into three categories. First there are those for which θ is rational, which are the farthest from being random. Then there are the θ irrational, but with excellent rational approximations, so that even though $(n\theta)$ is u.d. mod 1, it does not have uniformly low discrepancy. Finally, there are the irrational θ with poor rational approximations, and correspondingly low discrepancy in the $(n\theta)$ sequence.

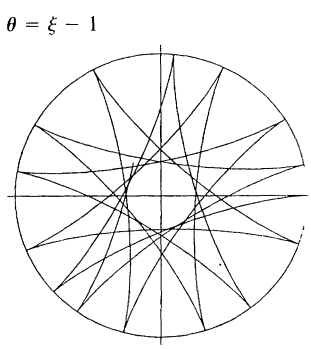


FIG. 6.

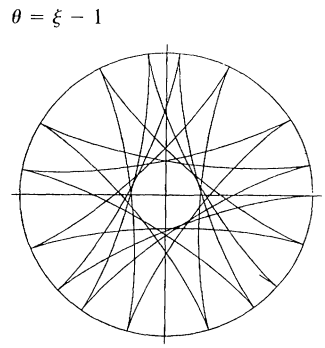


FIG. 7.

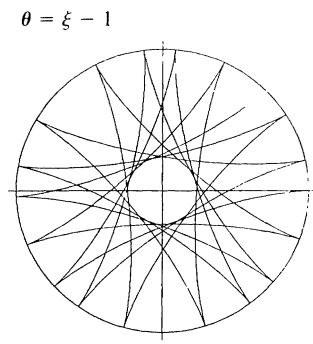


FIG. 8.

As a footnote, it should be recalled that $\xi = (1 + \sqrt{5})/2$, sometimes called the "golden ratio," is closely connected with the Fibonacci sequence. One of the many appearances of the Fibonacci numbers in nature has to do with the phenomenon known as *spiral phyllotaxis*, exemplified by the helical arrangement of leaves on many stemmed plants [1]. It turns out that in many species the angle between consecutive leaf buds is relatively constant at $2\pi(\xi - 1)$. Hence, the leaf buds,

viewed as points on a circle when looking down the stem, are the cusp points of the $\xi - 1$ hypocycloid. The result is that the number of spirals required for leaves to roughly line up on the stem is a Fibonacci number, and the number of leaves appearing in that collection of spirals is a Fibonacci number. Although the mechanisms are not well understood, it could be argued that the resulting leaf pattern, by avoiding clumping to the optimal extent, maximizes the exposure of leaves to the sun.

REFERENCES

1. H. S. M. Coxeter, Introduction to Geometry, John Wiley & Sons, Inc., New York, 1961.
2. Donald E. Knuth, The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, Addison-Wesley, Reading, Mass., 1969.
3. J. F. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1., *Mathematica B*, (Zutphen) 11 (1942/43) 11–17.
4. L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences, John Wiley & Sons, Inc., New York, 1974.
5. William J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, Mass., 1977.
6. H. Niederreiter, Discrepancy and convex programming, *Ann. Mat. pura Appl.*, (IV) 93 (1972) 89–97.

The Perimeter of a Rose*

JAMES P. BUTLER
Department of Enviromental Science and Physiology, Harvard School of Public Health, Boston, MA 02115

Introduction. Piranian [5] posed some interesting questions regarding the length $L_n(t)$ of the lemniscate level set

$$E(Q_n, t) = \{z: |Q_n(z)| = t\},$$

where $Q_n(z) = z^n - 1$, $n = 1, 2, \ldots$, and $0 \leqslant t < \infty$. (The case $n = 1$ is trivial; $L_1(t) = 2\pi t$. In what follows, take $n \geqslant 2$.) $E(Q_n, 1)$ is a rosette of n leaves; $E(Q_n, t)$ is a disjoint or smoothed version of the rosette for $t < 1$ and $t > 1$ respectively. FIG. 1 shows E for a few values of n and t. The basic questions concern whether (or where) $L_n(t)$ is an increasing function of t. Piranian gives elementary arguments to show several things. First, there are only finitely many indices n such that $dL_n(t)/dt = L_n'(t) > 0$ on $0 \leqslant t < \infty$. (He shows that $L_n(2) < L_n(1)$ for $n \geqslant 4$.) Second, $L_n'(t) > 0$ for all t in the interval $0 \leqslant t < 1$. He suggests that $L_3(t) < L_3(1)$ for some $t > 1$ (true) and that perhaps $L_2'(t) = 0$ at $t = 1$ (false). He poses three questions, all of which we answer in this note (including the case $n = 2$).

(a) Do the left- and right-hand derivatives of $L_n(t)$ exist at $t = 1$, and are they equal? Answer: they are respectively $+\infty$ and $-\infty$ for all $n \geqslant 2$.
(b) Is it true that for $n = 2, 3, \ldots$, the function $L_n(t)$ decreases in some interval $1 < t < T_n$ and increases in $T_n < t$? Answer: yes (note that $n = 2$ is included here).
(c) What is the asymptotic behavior of T_n, as $n \to \infty$? Answer: $T_n \sim \sqrt{n/2}$.

*Supported in part by NIH grant HL33009.

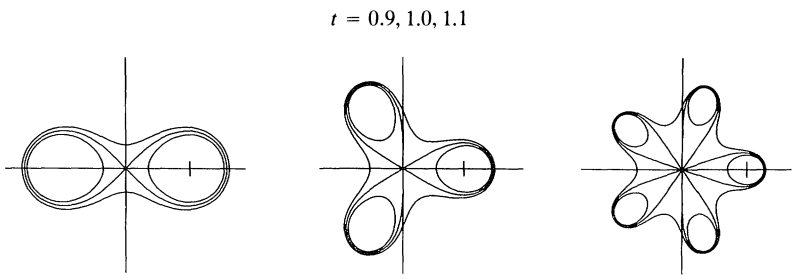


FIG. 1.

Explicit solution for $L_n(t)$, $t \geqslant 1$. The equation $|Q_n(z)| = t$ becomes, with $z = re^{i\theta}$,

$$r^{2n} - 2r^n \cos(n\theta) - u = 0,$$

where $u = t^2 - 1$. In the formulas which follow, the point $t = 1$ is understood to be the limit as $t \to 1^+$. The differential arc length $ds = d\theta(r^2 + (dr/d\theta)^2)^{1/2}$, and so, solving the above equation for r and replacing θ by $\xi = n\theta$ (to count the n leaves separately), we have

$$L_n(t) = 2\int_0^{\pi} d\xi \left(\frac{1+u}{\cos^2\xi + u}\right)^{1/2} \left[\cos\xi + (\cos^2\xi + u)^{1/2}\right]^{1/n}, \qquad u = t^2 - 1.$$

Now, being mindful of the sign of the cosine, write $x = \pm\cos\xi$ on $(0, \pi/2)$ and $(\pi/2, \pi)$, respectively. This yields

$$L_n(t) = 2t\int_0^1 dx \frac{\left[(x^2+u)^{1/2} + x\right]^{1/n} + \left[(x^2+u)^{1/2} - x\right]^{1/n}}{\sqrt{(1-x^2)(x^2+u)}}, \qquad u = t^2 - 1.$$

The transformation $w = 2x^2/u$ yields

$$L_n(t) = tw_0^{-(1-n)/2n} \int_0^{w_0} dw \frac{\left(\sqrt{w+2} + \sqrt{w}\right)^{1/n} + \left(\sqrt{w+2} - \sqrt{w}\right)^{1/n}}{\sqrt{w(w_0 - w)(w+2)}},$$

$$w_0 = \frac{2}{t^2 - 1}.$$

A slightly generalized form of this integral appears in Gradshteyn [**4**] as formula 3.277.7, who obtained it from Bateman [**2**], formula II: 13.1(12). Unfortunately, it was incorrectly copied; unfortunately, there is also a typographical error in Bateman's formula which propagated as well. The correct formula can be derived from Bateman [**3**], formula I: 3.7(8), and is

$$\int_0^y dx \frac{(y-x)^{\mu-1}}{\sqrt{x(x+2)}} \left[\left(\sqrt{x+2} + \sqrt{x}\right)^{2\nu} + \left(\sqrt{x+2} - \sqrt{x}\right)^{2\nu}\right]$$

$$= \sqrt{\pi}\,\Gamma(\mu) 2^{(2\nu+1)/2} \left[y(y+2)\right]^{(2\mu-1)/4} P_{\nu-1/2}^{1/2-\mu}(y+1),$$

where $P_a^b(z)$ is the associated Legendre function of the first kind, of degree a, order b, and argument z. Our case corresponds to $\mu = 1/2$, so we thus obtain the explicit solution for $L_n(t)$, $t \geqslant 1$ in terms of Legendre functions:

$$L_n(t) = 2\pi u^{\beta}\sqrt{1+u}\,P_{\beta}\left(1+\frac{2}{u}\right), \qquad 0 \leqslant u = t^2 - 1 < \infty, \qquad \beta = \frac{1-n}{2n}$$

$$= 2\pi t(t^2-1)^{(1-n)/2n} P_{(1-n)/2n}\left(\frac{t^2+1}{t^2-1}\right), \qquad t \geqslant 1.$$

A referee pointed out that this can also be conveniently expressed in terms of the Gauss hypergeometric function. In particular,

$$L_n(t) = 2\pi t^{1/n} F(-\beta, -\beta; 1; 1/t^2), \qquad \beta = \frac{1-n}{2n}, \qquad t \geqslant 1.$$

Explicit solution for $L_n(t)$, $0 \leqslant t \leqslant 1$. The procedure here is identical to the case $t > 1$, except that some care must be taken in the limits of the ξ integral. (Here the point $t = 1$ is the limit as $t \to 1^-$.) For the smoothed rosette above, ξ had limits 0 to π (which we divided into $(0, \pi/2)$ and $(\pi/2, \pi)$), and $dr/d\xi \leqslant 0$ in $(0, \pi)$. In the disjoint case $t < 1$, ξ has limits $(0, \xi_0)$, where $\xi_0 = \cos^{-1}(\sqrt{1-t^2})$. For ξ in this range, there is an outer and an inner portion of the curve, for which $dr/d\xi < 0$ and $dr/d\xi > 0$, respectively. Summing these yields the same types of formulas as those given above, with some sign changes. The result is

$$L_n(t) = 2\pi\sqrt{1+u}\,(-u)^{\beta} P_{\beta}\left(-1-\frac{2}{u}\right), \quad -1 \leqslant u = t^2 - 1 \leqslant 0, \quad \beta = \frac{1-n}{2n}$$

$$= 2\pi t(1-t^2)^{(1-n)/2n} P_{(1-n)/2n}\left(\frac{1+t^2}{1-t^2}\right), \qquad 0 \leqslant t \leqslant 1.$$

In terms of the hypergeometric function, this is equivalent to

$$L_n(t) = 2\pi t F(-\beta, -\beta; 1; t^2), \qquad \beta = \frac{1-n}{2n}, \qquad t \leqslant 1.$$

We may combine the above results by writing our final answer in the form

$$L_n(t) = 2\pi t|t^2-1|^{(1-n)/2n} P_{(1-n)/2n}\left(\frac{t^2+1}{|t^2-1|}\right), \qquad 0 \leqslant t < \infty.$$

This function is shown in FIG. 2 for $n = 2$, 3, and 5. Note the singularity at $t = 1$, and, near $t = 1.3$, the curious near coincidence of the curves.

Limits. We distinguish here first the limits as $t \to 0, 1^-, 1^+, \infty$, for n fixed, and second, the limit as $n \to \infty$ for fixed $t < 1$, $= 1$, and > 1. For n fixed, the limits $t \to 0$ and $t \to \infty$ correspond to the Legendre argument approaching 1^+, for which $P \to 1$ and $L_n(t) \sim 2\pi t$ and $2\pi t^{1/n}$, respectively. For n fixed and $t \to 1^-$ or 1^+, the argument approaches $+\infty$, and we use the leading behavior of $P_{\beta}(z)$ as $z \to \infty$:

$$P_{\beta}(z) \sim 2^{-\beta}\frac{\Gamma(1+2\beta)}{\Gamma^2(1+\beta)} z^{\beta} + 2^{\beta+1}\frac{\Gamma(-1-2\beta)}{\Gamma^2(-\beta)} z^{-\beta-1}, \qquad z \to \infty.$$

For fixed n at $t = 1$, only the first term survives in the expression for $L_n(1)$, while the second term contributes to the approach to $L_n(1)$.

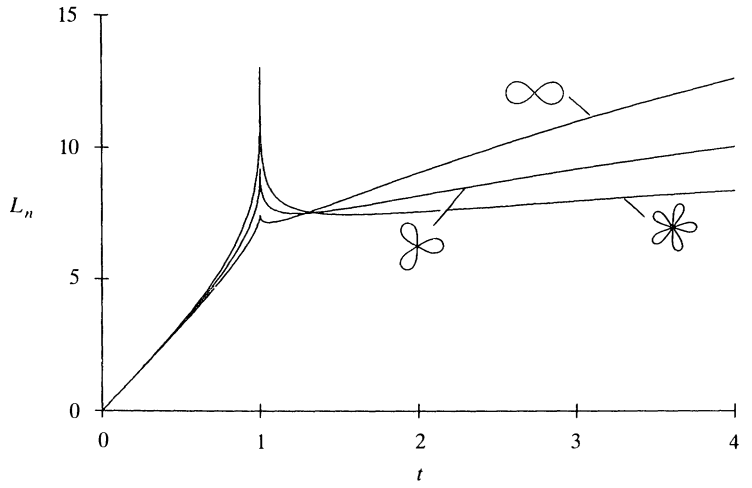


FIG. 2.

The second set of limits to consider are those for $n \to \infty$ with t fixed. For $t = 1$, the result above suffices, and we get the obvious limiting behavior $L_n(1) \sim 2n$. For fixed $t \neq 1$, the limit $n \to \infty$ corresponds to the Legendre degree approaching $-1/2$, for which we need the formula (1) $P_{-1/2}(\cosh \eta) = (2/\pi)K(\tanh(\eta/2))/\cosh(\eta/2)$, where K is the complete elliptic integral of the first kind, with modulus (not parameter) given as its argument. Using these limiting forms and the duplication formula for the gamma function in the explicit solutions given above, we may summarize these limits as follows (limit and asymptotic behavior are indicated with $\to$ and $\sim$, respectively):

n	t	$L_n(t)$
fixed	$\to 0^+$	$\sim 2\pi t$
fixed	$\to 1^{\pm}$	$\sim \sqrt{\pi}\left(2^{1/n}\dfrac{\Gamma\left(\frac{1}{2n}\right)}{\Gamma\left(\frac{1}{2n}+\frac{1}{2}\right)}+\dfrac{\Gamma\left(-\frac{1}{2n}\right)}{\Gamma\left(-\frac{1}{2n}+\frac{1}{2}\right)}\lvert t-1\rvert^{1/n}\right)$
fixed	$\to \infty$	$\sim 2\pi t^{1/n}$
$\to \infty$	fixed, < 1	$\to 4tK(t)$
$\to \infty$	fixed, $= 1$	$\sim 2n$
$\to \infty$	fixed, > 1	$\to 4K(1/t)$

The second entry answers question (a) of the introduction, and part of question (b). That is, the exponent of $|t-1|$ describing the approach to $L_n(1)$ is $1/n$; furthermore, $\Gamma(-1/2n)/\Gamma(-1/2n+1/2)$ is negative for $2 \leqslant n$, and thus $dL_n(t)/dt$ approaches $+\infty$ for $t \to 1^-$ and $-\infty$ for $t \to 1^+$. Except for the trivial case $n = 1$, $L_n(t)$ exhibits an infinitely sharp (symmetric) cusp at $t = 1$.

Where is $dL_n/dt = 0$? For $1 \leqslant t$, we know that $L_n(t)$ decreases up to some point T_n. Can we find an expression for T_n? It is straightforward to differentiate

the expression given above for $L_n(t)$, $t > 1$. Setting $dL_n(t)/dt = 0$, and using the relations among the contiguous Legendre functions to eliminate $dP_\beta(z)/dz$, we get a very pretty formula satisfied by T_n:

$$(1 + \beta)P_\beta(z) + \beta P_{\beta-1}(z) = 0, \qquad z = \frac{T_n^2 + 1}{T_n^2 - 1}, \qquad \beta = \frac{1 - n}{2n}.$$

Now as $n \to \infty$, T_n increases, and so we may examine this expression in the vicinity of $z = 1^+$. For this, we use the first two terms in the expansion of P_β:

$$P_\beta(z) = 1 + \frac{1}{2}\beta(\beta + 1)(z - 1) + \cdots, \qquad z \to 1^+.$$

The resulting solution is the leading term in the asymptotic behavior of T_n:

$$T_n \sim \sqrt{n/2}\,.$$

This is the answer to question (c).

The second half of question (b), equivalent to T_n being unique for each n, is answered by showing that the function $\Lambda_n(x) = L_n(t)$, $x = t^{1/n}$ is convex in $1 \leqslant x < \infty$. This is straightforward, though a bit tedious. We will not prove this here.

Some heuristics. Many of these results are satisfying intuitively. Consider t in the vicinity of 1, and examine the level curves of $Q_n(z)$ in the complex plane. As t moves from 1^- through 1 to 1^+, nothing very exciting happens to the curves except near the origin. As t passes through 1, we find curves lying in every other valley sharply descending to the valleys' bottoms and "popping" through to the complementary set of valleys. We call such a pop an nth order pop, the strength of which is given by the exponent of the cusp, namely $1/n$.

We conclude by giving a recipe for the behavior of the length of the level set $E(A_n, t)$ corresponding to an arbitrary nth order polynomial $A_n(z)$ of the form $\prod_{j=1}^n (z - z_j)$, where for convenience we will take the roots z_j distinct. Let $d_{ij} = |z_i - z_j|$, and let $D_i = \prod_{j \neq i} d_{ij}$. Draw the magnitude contour map of the given polynomial in the complex plane. For small t, the contours will be small circles surrounding the zeroes of A. In this limit, it is easy to see that $L \sim 2\pi t \sum_{i=1}^n D_i^{-1}$. For large t, the contour will approach a single circle, and, trivially, $L \sim 2\pi t^{1/n}$. For intermediate values of t, examine the contour map. As the small t circles increase in size, they will bump into each other with pops of varying order. At these values of t, L will exhibit cusps of the appropriate order.

Acknowledgements. I thank Mlle. H. Peslin for great help in performing the delicate numerical calculations to check this work, and in producing the graphics.

REFERENCES

1. M. Abramowitz and I. A. Stegun, eds., Handbook of Mathematical Functions, National Bureau of Standards, Washington, D.C., 1964.
2. A. Erdelyi, ed., Tables of Integral Transforms, Vol. II, Bateman Manuscript Project, New York, McGraw-Hill, 1954.
3. _____, ed. Higher Transcendental Functions, vol. I, Bateman Manuscript Project, New York, McGraw-Hill, 1953.
4. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, fourth ed., Academic Press, New York, 1965.
5. G. Piranian, The length of a lemniscate, this MONTHLY, (87) 555–556, 1980.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A Tree Among the Stepping Stones

A. R. MEIJER
Department of Mathematics and Applied Mathematics, University of Natal, Durban 4001, South Africa

The transportation problem is a linear programming (LP) problem of the form:

$$\text{minimize } \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} x_{ij}$$

subject to the constraints

$$\sum_{j=1}^{n} x_{ij} = a_i \quad \text{for } i = 1, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} = b_j \quad \text{for } j = 1, \ldots, n,$$

and $x_{ij} \geqslant 0$ for all i and j.

The traditional interpretation of this problem is that of m factories with outputs $a_1, \ldots, a_m$ and n warehouses with requirements $b_1, \ldots, b_n$, and with c_{ij} denoting the cost of moving one unit from factory i to warehouse j, the goal being to minimize the total cost of transporting the goods from factories to warehouses. Note that, since

$$\sum_i a_i = \sum_i \sum_j x_{ij} = \sum_j b_j,$$

the x_{ij} must satisfy $m + n - 1$ independent equations, so any basic feasible solution will contain $m + n - 1$ basic variables.

Because of the special structure of this LP problem, special techniques for solving it have evolved, popular among them being the so-called stepping stone method [1]. The essence of this method lies in the following fact: If a basic feasible solution has been obtained, then for any nonbasic variable x_{ij}, there exists a (unique) "loop"

$$x_{ij}, x_{ij_1}, x_{i_1 j_1}, x_{i_1 j_2}, x_{i_2 j_2}, \ldots, x_{i_{2k-1} j_{2k-1}}, x_{i_{2k-1} j}, x_{ij}$$

beginning and ending at x_{ij} and alternately changing rows and columns, in which all the other variables are currently basic. For each nonbasic variable x_{ij} one may now calculate the net cost involved in increasing x_{ij} by one unit, which necessitates decreasing x_{ij_1} by one unit, which necessitates increasing $x_{i_j j_1}$ by one unit, and so on. If this net cost is negative, that is, if it is advantageous to do so, we incorporate x_{ij} into the basis and give it the largest value possible without violating the nonnegativity constraints, which implies that (at least) one of the basic variables $x_{ij_{l+1}}$ becomes zero and this variable (arbitrarily chosen if there is more than one) is now dropped from the basis.

Most of the standard treatments of the stepping stone algorithm, such as [4, 5, 7], do not make clear why such a "loop" should necessarily exist, nor why it should be unique. It is interesting to note that the authors of [1] missed the uniqueness property, advising the reader that "when alternate stones are available along the route he should take the shortest path from the beginning row to the ending column" [1, p. 54]. Existence and uniqueness are proved in [6, p. 385] by means of a linear algebraic argument. The purpose of this note is to sketch a transparent graph-theoretical proof, which depends only on the following two properties of trees (see, e.g. [2, p. 69]):

1. A connected graph on ν vertices is a tree if and only if the graph has exactly $\nu - 1$ edges;
2. In a tree, given any two vertices, there is a unique path connecting them.

With the increasing emphasis on discrete mathematics, many, if not all, students are familiar with these ideas.

If a basic solution to the transportation problem is given, we consider the graph with vertex set $\{r_1, r_2, \ldots, r_m, c_1, c_2, \ldots, c_n\}$, where r_k and c_l are connected by an edge if and only if x_{kl} is a basic variable. The loop above then corresponds to a path from r_i to c_j. The required results will have been proved once it has been shown that the graph is a tree. (The latter fact is equivalent to Dantzig's Basis Triangularity theorem [3, p. 303]). Since the graph has $m + n$ vertices and $m + n - 1$ edges, it is sufficient to prove that it is connected.

Note, however, that any iteration of the stepping stone algorithm preserves connectedness, since if x_{ij} is taken into the basis (that is, if r_i and c_j are connected by an edge), a cycle is completed in the graph, this cycle then being removed by the deletion of $x_{i_l j_{l+1}}$ from the basis (that is, by the removal of the edge from r_{i_l} to $c_{j_{l+1}}$). It therefore suffices to show that the initial basic feasible solution will indeed yield a tree.

We shall discuss this only for the simplest method of finding a starting basic solution, namely the so-called "north-west corner" method. In this method one starts by choosing x_{11} as a basic variable and giving it the value $\min(a_1, b_1)$. If $x_{11} = a_1$, all the variables $x_{1j} (j = 2, \ldots, n)$ are declared to be nonbasic; if $x_{11} = b_1$, the $x_{i1} (i = 1, \ldots, m)$ are designated nonbasic. If $a_1 = b_1$, one, *but not both*, of the sets $\{x_{12}, \ldots, x_{1n}\}$ and $\{x_{21}, \ldots, x_{m1}\}$ is designated to consist of nonbasic variables. The subgraph $\{r_1, c_1\}$ is now connected, obviously.

At any subsequent stage, choose x_{ij} from among the variables that have not yet been designated basic or nonbasic and with either i or j minimal among all the indices appearing in this set. Designate this x_{ij} as basic, giving it the maximum value possible within the constraints and the values already given to the basic variables selected earlier. (This value may have to be zero, but this degeneracy does not cause problems.) As before, all the remaining variables in either the ith row or the jth column (but not both) are designated nonbasic. From our choice of x_{ij}, x_{ik} has already been selected as basic for some $k < j$, or else x_{lj} has, for some $l < i$. Thus we connect r_i and c_j where either (r_i, c_k) or (r_l, c_j) was a previously existing edge and (by the induction hypothesis) the graph previously constructed was connected. Because, from the constraints, every row and column must contain a basic variable once the north-west corner procedure has been concluded (that is,

once $m + n - 1$ variables have been designated basic), this completes the argument.

The author thanks the referees for their constructive suggestions.

REFERENCES

1. A. Charnes and W. Cooper, The stepping stone method of explaining linear programming in transportation problems, *Management Sci.*, 1(1954) 46–49.
2. G. Chartrand and L. Lesniak, Graphs and Digraphs, 2nd ed., Wadsworth, Monterey, 1986.
3. G. B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, 1963.
4. L. R. Foulds, Optimization Techniques, Springer, New York, 1981.
5. C. M. Harvey, Operations Research, North-Holland, New York, 1979.
6. K. G. Murty, Linear Programming, Wiley, New York, 1983.
7. H. A. Taha, Operations Research—an Introduction, Macmillan, New York, 1987.

An Analytic Technique to Prove Borel's Strong Law of Large Numbers

LIU WEN
Department of Mathematics, Hebei Institute of Technology, Tianjin 300130,People's Republic of China

Borel's strong law of large numbers asserts that if S_n, $n \geq 1$, is the number of successes in the first n independent repetitions of a Bernoulli trial with success probability p, $0 < p < 1$, then

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1. \tag{1}$$

This result is commonly proved in textbooks through the application of powerful probabilistic tools or by a combinatorial argument (see, e.g. [**1**, p. 42], [**2**, p. 97], or [**4**, p. 273]). Tomkins [**6**] and Taylor and Hu [**5**] provided additional proofs which, although simpler than the proofs common in most textbooks, use the same basic probabilistic results. In this note we present a new analytical technique to prove Borel's strong law of large numbers, using tools different than is usual. The crucial part is the application of Lebesgue's theorem on differentiability of monotone functions (see [**3**, p. 358]) to the study of a.e. convergence.

We first give a representation of a set of independent and identically distributed Bernoulli random variables with success probability p $(0 < p < 1)$ in the probability space $([0, 1), \mathscr{B}, \lambda)$, where $\mathscr{B}$ denotes the Borel sets and λ is Lebesgue measure. Divide $[0, 1)$ into two right-semi-open intervals: $d_0 = [0, 1 - p)$, $d_1 = [1 - p, 1)$. These intervals will be called intervals of first order. Proceeding inductively, suppose the 2^n nth-order intervals $d_{x_1 \cdots x_n}$ $(x_i = 0, 1;\ i = 1, 2, \ldots, n)$ have been defined. Then divide each interval $d_{x_1 \cdots x_n}$ into two right-semi-open intervals $d_{x_1 \cdots x_n 0}$ and $d_{x_1 \cdots x_n 1}$ according to the ratio $(1 - p) : p$; this creates 2^{n+1} intervals of order $n + 1$. Define, for $n \geq 1$, random variables X_n: $[0, 1) \to \{0, 1\}$ as follows:

$X_n(x) = x_n$ if $x \in d_{x_1 \cdots x_n}$. It is easy to see that

$$\lambda\left(d_{X_1(x)X_2(x)\cdots X_n(x)}\right) = p^{S_n(x)}(1-p)^{n-S_n(x)}, \tag{2}$$

where $S_n(x) = X_1(x) + \cdots + X_n(x)$. It follows from (2) that the X_n, $n \geq 1$, are independent and identically distributed Bernoulli random variables with $\text{Prob}(X_n = 1) = p$ and $\text{Prob}(X_n = 0) = 1 - p$.

In order to prove (1) we first construct an auxiliary function. Let $r \in (0,1)$ be a constant. Dividing the interval $[0,1)$ successively according to the ratio $(1-r):r$ as before, another set of nth-order intervals $D_{x_1 \cdots x_n}$, $(x_i = 0,1;\ i = 1,2,\ldots,n)$ is created. Let $d^-_{x_1 \cdots x_n}$ and $d^+_{x_1 \cdots x_n}$ be, respectively, the left and right endpoints of $d_{x_1 \cdots x_n}$; define $D^-_{x_1 \cdots x_n}$ and $D^+_{x_1 \cdots x_n}$ similarly. Let Q be the set of endpoints of all d-intervals.

Now we define an increasing function $f_r\colon [0,1) \to [0,1)$ as follows:

$$\begin{aligned} f_r\left(d^-_{x_1 \cdots x_n}\right) &= D^-_{x_1 \cdots x_n}; \\ f_r\left(d^+_{x_1 \cdots x_n}\right) &= D^+_{x_1 \cdots x_n}; \\ f_r(x) &= \sup\{f(t)\colon t \in Q \cap [0,x)\}, \quad \text{if } x \in [0,1) \setminus Q. \end{aligned}$$

Let

$$t_n(r,x) = \frac{f_r\left(d^+_{x_1 \cdots x_n}\right) - f_r\left(d^-_{x_1 \cdots x_n}\right)}{d^+_{x_1 \cdots x_n} - d^-_{x_1 \cdots x_n}}$$

for $x \in d_{x_1 \cdots x_n}$. It is easy to see that

$$\begin{aligned} t_n(r,x) &= \frac{\lambda\left(D_{X_1(x)\cdots X_n(x)}\right)}{\lambda\left(d_{X_1(x)\cdots X_n(x)}\right)} = \frac{r^{S_n(x)}(1-r)^{n-S_n(x)}}{p^{S_n(x)}(1-p)^{n-S_n(x)}} \\ &= \left(\frac{r(1-p)}{p(1-r)}\right)^{S_n(x)} \left(\frac{(1-r)}{(1-p)}\right)^n. \end{aligned} \tag{3}$$

Letting $A(r)$ be the set of points of differentiability of f_r, we have $\lambda(A(r)) = 1$ by Lebesgue's theorem on differentiability of monotone functions. By a property of the derivative (see [3, p. 345]) we have $\lim_{n\to\infty} t_n(r,x) = f_r'(x) < \infty$ for $x \in A(r)$. Thus

$$\limsup_{n\to\infty} \frac{\log t_n(r,x)}{n} \leq 0, \quad \text{for } x \in A(r), \tag{4}$$

where log denotes natural logarithm. By (3) and (4),

$$\limsup_{n\to\infty} \frac{S_n(x)}{n} \log \frac{r(1-p)}{p(1-r)} \leq \log \frac{1-p}{1-r}. \tag{5}$$

If $p < r < 1$, we have by (5) that

$$\limsup_{n\to\infty} \frac{S_n(x)}{n} \leq \frac{\log \dfrac{1-p}{1-r}}{\log \dfrac{r(1-p)}{p(1-r)}}. \tag{6}$$

Assuming $p < r_k < 1$, $r_k \to p$ (as $k \to \infty$), and letting $A = \bigcap_{k=1}^{\infty} A(r_k)$, for each $k = 1, 2, \ldots$, we have, by (6),

$$\limsup_{n\to\infty} \frac{S_n(x)}{n} \le \frac{\log \dfrac{1-p}{1-r_k}}{\log \dfrac{r_k(1-p)}{p(1-r_k)}}, \quad \text{for } x \in A. \tag{7}$$

Since

$$\lim_{k\to\infty} \frac{\log \dfrac{1-p}{1-r_k}}{\log \dfrac{r_k(1-p)}{p(1-r_k)}} = p,$$

(7) implies that

$$\limsup_{n\to\infty} \frac{S_n(x)}{n} \le p, \quad \text{for } x \in A. \tag{8}$$

Similarly, assuming $0 < l_k < p$, $l_k \to p$ (as $k \to \infty$), and letting $B = \bigcap_{k=1}^{\infty} A(l_k)$, we have

$$\liminf_{n\to\infty} \frac{S_n(x)}{n} \ge p, \quad \text{for } x \in B. \tag{9}$$

From (8) and (9) it follows that

$$\lim_{n\to\infty} \frac{S_n(x)}{n} = p, \quad \text{for } x \in A \cap B. \tag{10}$$

Because $\lambda(A \cap B) = 1$, (1) follows from (10).

REFERENCES

1. Y. S. Chow and H. Teicher, Probability Theory: Independence, Interchangeability, Martingales, Springer-Verlag, New York, 1978.
2. K. L. Chung, A Course in Probability Theory, Harcourt, New York, 1968.
3. T. H. Hildebrandt, Introduction to the Theory of Integration, Academic Press, New York, 1963.
4. V. K. Rohatgi, An Introduction to Probability Theory and Mathematical Statistics, Wiley, New York, 1976.
5. R. L. Taylor and T. C. Hu, Sub-Gaussian techniques in proving strong laws of large numbers, this MONTHLY, 94 (1987) 295–299.
6. R. J. Tomkins, Another proof of Borel's strong law of large numbers, *Amer. Statist.*, 38 (1984) 208–209.

A Topological Mean Value Theorem for the Plane

IRA ROSENHOLTZ
Department of Mathematics, Eastern Illinois University, Charleston, IL 61920

Cauchy's Generalized Mean Value Theorem states that if x and y are functions that are continuous on $[a, b]$ and differentiable on (a, b), then there is a t_0 in (a, b) such that $x'(t_0)[y(b) - y(a)] = y'(t_0)[x(b) - x(a)]$. One geometric interpretation of this is that if $\Omega(t) = (x(t), y(t))$ is a path in the plane, then, under certain hypotheses, there must be an instant at which the velocity vector is *parallel* to the vector joining the endpoints of the path. Can this conclusion be improved? In particular, can we get a velocity vector to point in the *same direction* as the vector from the first endpoint to the last and not just be parallel? This is the question we address in this note. Surprisingly, the answer turns out to be a consequence of the Jordan Curve Theorem! This result provides some additional insight into such old standards as the Mean Value Theorem and Cauchy's Generalization, l'Hôpital's Rule, and Darboux's Theorem, while demonstrating further interplay between topology and analysis.

Consider the related question: *Can one go north without ever heading north*? A little reflection shows that the answer to this question is *Yes* and, in fact, this can happen in a variety of ways. First of all, one might travel up to the North Pole on a globe by "spiraling up" to it (see FIGURE 1a). (Of course, it is even easier to go east by heading west, but that's another story.) Even if traveling is restricted to the plane, it can still be done. One way is to stop (i.e., have zero velocity vector at some instant). Then one's path can have a "differentiable corner" (see FIGURE 1b). And another way to go north without ever heading north is to head SOUTH! (See FIGURE 1c.) Notice that in this example the path crosses itself. As we shall soon see, these are the only things that can go wrong.

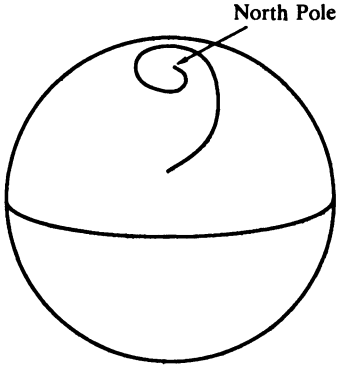


FIG. 1(a)

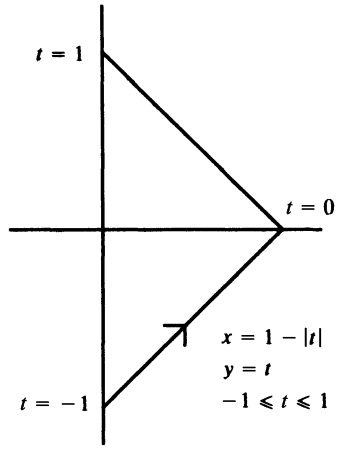


FIG. 1(b)

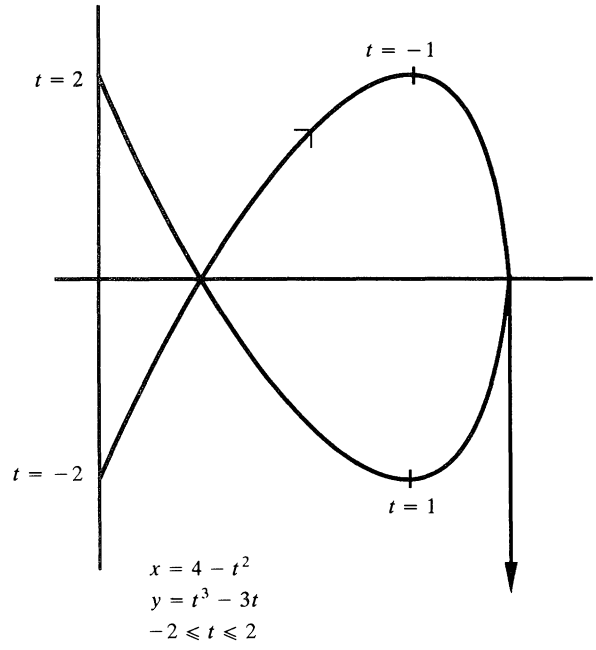


FIG. 1(c)

Let us agree on the following vocabulary: A *path* in the plane is a continuous function $\Omega(t) = (x(t), y(t))$ defined on some closed interval $[a, b]$ of real numbers, where $a < b$. A path Ω is *differentiable* provided $\Omega'(t) = (x'(t), y'(t))$ exists for $a < t < b$. (We do not assume that Ω is continuously differentiable.) A differentiable path is *nonstop* if $\Omega'(t) \neq (0,0)$ for all t such that $a < t < b$. Finally, a path that is one-to-one is called an *arc*.

We prove the following:

TOPOLOGICAL MEAN VALUE THEOREM. *If* Ω: $[a, b] \to \mathbb{R}^2$ *is a differentiable, nonstop arc in the plane, then there is a positive number* μ *and a* t_0 *in* (a, b) *such that* $\Omega'(t_0) = \mu[\Omega(b) - \Omega(a)]$. *That is, at some instant* t_0, *the velocity vector* $\Omega'(t_0)$ *points in the same direction as the vector from* $\Omega(a)$ *to* $\Omega(b)$, *or, what amounts to the same thing,* $\Omega'(t_0)/\|\Omega'(t_0)\| = [\Omega(b) - \Omega(a)]/\|\Omega(b) - \Omega(a)\|$.

(Notice that this generalizes the Mean Value Theorem: let $\Omega(t) = (t, f(t))$ on $[a, b]$.)

Proof of the Theorem. First note that it suffices to consider the special case where $\Omega(b)$ is due north of $\Omega(a)$ and both lie on the y-axis. (If $\Omega(b)$ is not due north of $\Omega(a)$, consider the auxiliary path $\beta(t) = (0, 1)[\Omega(t) - \Omega(a)]/[\Omega(b) - \Omega(a)]$, where all points in the plane are to be thought of as complex numbers, and all operations are as such. Then β is also a differentiable nonstop arc in the plane, and $\beta(b) = (0, 1)$ is due north of $\beta(a) = (0, 0)$. Furthermore, the μ and t_0 that work for β also work for Ω.) Let t_1 in $[a, b]$ be chosen so that $\Omega(t_1)$ lies on the y-axis and is the most northerly such point. (It is possible that t_1 equals b—a situation we shall have to address—but certainly t_1 does not equal a.) There

are two possibilities: Case 1: There is an increasing sequence $s_1 < s_2 < \cdots$ such that $\lim s_n = t_1$ and $\Omega(s_n)$ lies on the y-axis for all n; and Case 2: There is no such sequence.

Case 1. We may assume in this case that t_1 is less than b. (If not, choose a c less than b so that $\Omega(c)$ is due north of $\Omega(a)$, and choose a new t_1 for the path Ω restricted to the interval $[a, c]$. Note that with this new choice of t_1, we might now be in Case 2.) So Ω is differentiable at t_1. Then $\Omega'(t_1) = \lim[\Omega(s_n) - \Omega(t_1)]/(s_n - t_1)$. But since $\Omega(s_n)$ lies on the y-axis below $\Omega(t_1)$, and $s_n < t_1$, we may conclude that $\Omega'(t_1)$ is of the form $(0, y'(t_1))$ with $y'(t_1) \geqslant 0$. Finally, since $\Omega'(t_1) \neq (0, 0)$ by hypothesis, $y'(t_1)$ must be greater than 0, and thus, at time t_1, the velocity vector is pointing due north as claimed.

Case 2. In this case, there is a $t_2 < t_1$ such that $\Omega(t_2)$ lies on the y-axis (below $\Omega(t_1)$, of course) and $\Omega(t)$ does not lie on the y-axis for all t between t_2 and t_1. Then for $t_2 < t < t_1$, Ω lies entirely to the left of the y-axis or entirely to the right. Without loss of generality assume the former, and concentrate on the restriction of Ω to the interval $[t_2, t_1]$. Since Ω is continuous on this interval, there is a t_0 such that $t_2 < t_0 < t_1$ and $\Omega(t_0)$ has minimal x-coordinate. Then $x'(t_0) = 0$. Therefore, since $\Omega'(t_0) \neq (0, 0)$, $\Omega'(t_0)$ points due north or due south. If $\Omega'(t_0)$ points north, we are done. So assume, to arrive at a contradiction, that $\Omega'(t_0)$ points south.

Now $\Omega([t_2, t_1])$ lies in the vertical strip $\{(x, y) | x(t_0) \leqslant x \leqslant 0\}$. Furthermore, since $\Omega'(t_0)$ points due south, then for all $t < t_0$ (resp., $t > t_0$) sufficiently close to t_0, $\Omega(t)$ lies above (resp., below) the line $y = y(t_0)$, and as close to vertical as you please. But $\Omega([t_2, t_0])$ separates the vertical strip, and in fact separates points nearly due south of $\Omega(t_0)$ from $\Omega(t_1)$. One easy way to see this is to extend $\Omega([t_2, t_0])$ to a simple closed curve by going due west from $\Omega(t_0)$, then going due north far enough to avoid the image of Ω, next going due east to get to the right of the y-axis, then proceeding south to the height of $\Omega(t_2)$, and finally due west to $\Omega(t_2)$ (see FIGURE 2). Points of $\Omega([t_0, t_1])$ almost due south of $\Omega(t_0)$ lie on the outside of this simple closed curve and $\Omega(t_1)$ lies on the inside of this simple closed curve. So by the Jordan Curve Theorem, $\Omega([t_0, t_1])$, which lies inside the strip, must cross this simple closed curve, necessarily at a point of $\Omega([t_2, t_0))$. This contradicts the hypothesis that Ω is an arc and, therefore, $\Omega'(t_0)$ points due north, as claimed, completing the proof. □

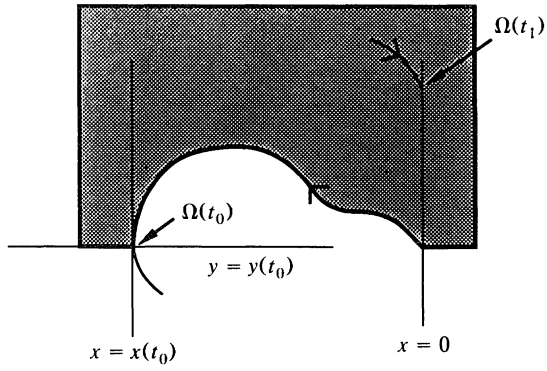


FIG. 2

We now derive several corollaries.

The first is about simple closed paths in the plane. (Here it may be convenient to think of a differentiable simple closed path in the plane as a nondegenerate periodic differentiable function from $\mathbb{R}$ to $\mathbb{R}^2$.) It says roughly that if you traverse a simple closed curve in a differentiable, nonstop fashion, then you must have been facing in every direction at least once. An amusing interpretation of this is that if a differentiable, nonstop closed path is *one-to-one* (i.e., simple), then the set of tangent directions (i.e., unit tangent vectors) is *onto* the unit circle. The reader is invited to compare this situation to closed paths that are either nonsimple, not nonstop, or nondifferentiable at even a single point.

COROLLARY 1. *If δ is a differentiable, nonstop simple closed path in the plane and u is any unit vector in the plane, then there is a t_0 such that $\delta'(t_0)/\|\delta'(t_0)\| = u$.*

The idea is to use the Jordan Curve Theorem again. Take a point in the inside of the simple closed curve and construct the line through this point parallel to u. This line intersects the simple closed curve in (at least) two points A and B. Part of the time our path goes from A to B, and part of the time it goes from B to A, and it is now a simple matter to apply the Topological Mean Value Theorem to obtain the result.

One of the pretty applications of Rolle's Theorem is the Intermediate Value Theorem for Derivatives (Darboux's Theorem): If f is differentiable on the closed interval $[a, b]$, then f' attains all values between $f'(a)$ and $f'(b)$. We prove the following analogue:

COROLLARY 2. (*Intermediate Value Theorem for Derivative Directions*). *Suppose that $\Omega\colon [a, b] \to R^2$ is a differentiable, nonstop arc in the plane. Suppose further that Ω has a nonzero derivative at both a and b. Then $\Omega'(t)/\|\Omega'(t)\|$ takes on every value on at least one of the two circular arcs between $\Omega'(a)/\|\Omega'(a)\|$ and $\Omega'(b)/\|\Omega'(b)\|$. And, therefore, the set of derivative directions forms a connected subset of the unit circle.*

(Recall that we are not assuming that Ω is continuously differentiable, or this wouldn't be very surprising. Also notice that this corollary implies Darboux's Theorem: let $\Omega(t) = (t, f(t))$ and use the fact that Ω' never points left.)

Proof. Consider the set $S = \{(s, t) | a \leqslant s < t \leqslant b\}$. Then S is connected, and the functions $h(s, t) = [\Omega(t) - \Omega(s)]/[t - s]$ and $H(s, t) = h(s, t)/\|h(s, t)\|$ are continuous on S. Therefore the image of H is a connected subset of the unit circle. And since it contains points arbitrarily close to $\Omega'(a)/\|\Omega'(a)\|$ and $\Omega'(b)/\|\Omega'(b)\|$, it must contain one of the two arcs of the unit circle between these two points. Now let A denote this circular arc. If u belongs to A, there is an (s, t) in S so that $H(s, t) = u$, i.e., so that $[\Omega(t) - \Omega(s)]/\|[\Omega(t) - \Omega(s)]\| = u$. We now apply our Topological Mean Value Theorem to the differentiable, nonstop arc Ω restricted to the interval $[s, t]$ to obtain a t_0 in (s, t) such that $\Omega'(t_0)/\|\Omega'(t_0)\| = u$, completing the proof.

Notice that we really do not require the full strength of the hypothesis that Ω be an arc. For example, this result can easily be generalized to the cases in which Ω is "locally one-to-one" and "piecewise one-to-one." We leave it to the interested reader to find his or her favorite generalization.

The statement of Corollary 2 suggests several natural questions. Suppose Ω is a differentiable, nonstop path which also has nonzero derivatives at its endpoints. *Must the set of velocity vectors be a connected subset of the plane*? If not, *must the set of speeds be a connected subset of the real line*? Finally, *must the set of derivative directions be a connected subset of the unit circle even if* Ω *is not piecewise one-to-one*? It turns out that the answer to all three questions is "No." Here is a monster-making "machine"*: Let β: $[1/2, 1] \to [1/2, 1] \times [-1/2, 1/2]$ be any continuously differentiable path such that:

(a) β is nonstop;
(b) $\beta(1/2) = (1/2, 0)$ and $\beta(1) = (1, 0)$;
(c) $\beta'(1/2) = \beta'(1)$;
(d) $\beta'(t) \neq (1, 0)$ for all t.

Roughly speaking, we extend β to a path Ω on the interval $[0, 1]$ which looks like β on $[1/2, 1]$ stuck to two copies of β of $1/4$ the dimensions on the interval $[1/4, 1/2]$ stuck to four copies of β of $1/16$ the dimensions on the interval $[1/8, 1/4]$, and so on, with $\Omega(0) = (0, 0)$ (see FIG. 3). The path Ω has as its velocity

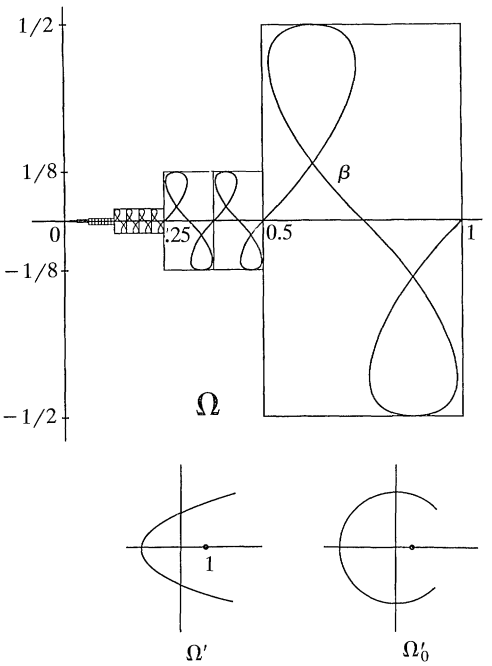


FIG. 3

*For readers who may be interested only in the answer to the first two questions, simpler examples exist. For example, if p, q, and r are real numbers, the path given by

$$\Omega(t) = \left(pt + rt^2\cos(1/t), qt + rt^2\sin(1/t)\right) \quad \text{if } 0 < t \leqslant 1 \text{ and } \Omega(0) = (0, 0)$$

has as its set of velocity vectors a spiral about a circle of radius r centered at (p, q), together with its center. However, our "machine" answers all three. In addition, the "machine" may be modified to produce a differentiable path whose set of velocity vectors is the entire plane.

vectors precisely the velocity vectors of β together with $\Omega'(0) = (1, 0)$. Finally, if β has length L, then Ω has length $2L$, and if we parametrize Ω using arc length then we obtain a path Ω_0 with $\Omega_0(0) = (0, 0), \Omega_0(2L) = (1, 0)$, and Ω_0 has unit speed except at time 0 at which time its speed is $1/2L$.

We conclude with a two-dimensional version of L'Hôpital's Rule. (Perhaps this is appropriate as L'Hôpital's Rule seems to be one of the more popular applications of Cauchy's Generalized Mean Value Theorem.)

COROLLARY 3. *Suppose that* $\Omega(t) = (x(t), y(t))$ *is a differentiable, nonstop path defined on the closed interval* $[a, b]$. *If* Ω *is one-to-one on some neighborhood of* a, *and* $\lim \Omega(t) = (0, 0)$ *as* t *approaches* a^+, *then as* t *approaches* a^+, $\lim \Omega(t)/\|\Omega(t)\| = \lim \Omega'(t)/\|\Omega'(t)\|$, *assuming the latter limit exists.*

One interpretation of this is that if a particle starts at the origin in the plane, then, under reasonable conditions, its "position directions" are closely approximated by its "velocity directions." Another interpretation, which is perhaps even more intuitive, is that if a particle has constant mass and zero initial velocity, then (under suitable conditions) its unit velocity vectors near the start are approximately its unit acceleration vectors (i.e., its force direction vectors). In other words, if a particle is at rest, then the direction it goes is pretty much the direction you push it. I guess the only real surprise here is that we are not even assuming that the force is continuous.

The proof of this corollary is identical to the usual proof of L'Hôpital's Rule given the Generalized Mean Value Theorem and is omitted.

Disks and Shells Revisited

WALTER CARLIP
Department of Mathematics, Ohio University, Athens, OH 45701

It is a common practice in calculus courses to use the definite integral to *define* the area between the graph of a function and the x-axis (see, e.g., [**2**, p. 252], [**3**, p. 221], and [**4**, p. 238]). Soon after, the student is taught two methods to calculate volumes of solids of revolution—the disk method and the shell method—usually with no mention of how *volume* is defined. Most calculus books follow the introduction of disks and shells with several examples in which it is shown that both methods of calculating the volume yield the same answer. The alert student is sure to wonder whether this is always the case.

The equivalence of the disk and shell methods was proven in [**1**] using integration by parts. We present here a different approach, one that uses only elementary ideas and illustrates an important proof technique.

THEOREM. *Let* $f(x)$ *be a continuous, invertible function on the interval* $[a, b]$, *where* $a \geqslant 0$. *Suppose the region bounded by* $y = f(b)$, $x = a$, *and the graph of* $f(x)$ *is rotated about the y-axis. Then the values of the volume of the resulting solid*

obtained by the disk and shell methods are equal. That is,

$$\int_{f(a)}^{f(b)} \left(\pi\left[f^{-1}(y)\right]^2 - \pi a^2\right) dy = \int_a^b 2\pi x[f(b) - f(x)]\, dx.$$

The Ingredients. Early in most calculus curricula, Rolle's Theorem is used to prove the following principle, which is then applied repeatedly.

PRINCIPLE. *If $f(x)$ and $g(x)$ are two functions that satisfy:*
(a) *$f(x)$ and $g(x)$ are differentiable on an interval $[a, b]$,*
(b) *$f'(x) = g'(x)$ for all $x \in [a, b]$, and*
(c) *$f(c) = g(c)$ for one point $c \in [a, b]$,*
then $f(x) = g(x)$ for all $x \in [a, b]$.

Informally, this says that two functions are equal if they are equal at *one* point and have identical derivatives. Although this principle is surprisingly simple, and easily absorbed by students, it has numerous applications and reappears often. Emphasizing this principle helps students see the similarity between proofs that otherwise seem unrelated.

The other ingredients of the proof of our theorem are the Fundamental Theorem of Calculus, the chain rule, and the product rule.

Proof. Let $t \in [a, b]$. We define two functions of t as follows:

$$V(t) = \int_{f(a)}^{f(t)} \left(\pi\left[f^{-1}(y)\right]^2 - \pi a^2\right) dy$$

and

$$W(t) = \int_a^t 2\pi x[f(t) - f(x)]\, dx.$$

We need to prove that $V(b) = W(b)$, as these are the two volumes in the theorem. It is easy to show that $V(t) = W(t)$ for all $t \in [a, b]$ by applying the principle given above.

First simplify $W(t)$:

$$W(t) = 2\pi f(t) \int_a^t x\, dx - 2\pi \int_a^t x f(x)\, dx.$$

Now, by the Fundamental Theorem, both $V(t)$ and $W(t)$ are differentiable on the interval $[a, b]$. Furthermore, $V'(t) = (\pi[f^{-1}(f(t))]^2 - \pi a^2)f'(t) = \pi(t^2 - a^2)f'(t)$, and

$$W'(t) = 2\pi f'(t) \int_a^t x\, dx + [2\pi f(t)t - 2\pi t f(t)]$$

$$= 2\pi f'(t)\left[\frac{t^2 - a^2}{2}\right] = \pi(t^2 - a^2)f'(t).$$

Thus, $W'(t) = V'(t)$ for all $t \in [a, b]$. It remains only to observe that $V(a) = W(a) = 0$, by a fundamental property of integrals. □

REFERENCES

1. Charles A. Cable, The disk and shell method, this MONTHLY, 91 (1984) 139.
2. John B. Fraleigh, Calculus of a Single Variable, Addison-Wesley, Reading, Mass., 1985.
3. Al Shenk, Calculus and Analytic Geometry, Goodyear Publ., Santa Monica, Calif., 1979.
4. Michael Spivak, Calculus, Publish or Perish Inc., Berkeley, Calif., 1980.

L'Hôpital's Rule Via Integration

DONALD HARTIG
Mathematics Department, California Polytechnic State University, San Luis Obispo, CA 93407

In elementary calculus texts L'Hôpital's rule is usually proven only for the case $0/0$, $x \to x_0$ (finite), by applying the Cauchy mean value theorem. Extension to $x \to \infty$ is then accomplished by replacing x with $1/x$. Verification of the rule for the ∞/∞ indeterminate form is regarded as too difficult and may be discussed in an exercise, an appendix, or not at all. In this note we give a proof for the ∞/∞ case that does not make use of the Cauchy mean value theorem. Instead, we require that the functions have continuous derivatives and take advantage of the order properties of the definite integral. The argument adapts nicely to the case $0/0$ as well.

L'HÔPITAL'S RULE. ∞ / ∞. *Let f and g have continuous derivatives with $g'(x) \neq 0$. If $\lim_{x\to\infty} f(x) = \infty$, $\lim_{x\to\infty} g(x) = \infty$, and $\lim_{x\to\infty} f'(x)/g'(x) = L$, then $\lim_{x\to\infty} f(x)/g(x) = L$ also.*

Proof. We assume that L is finite; the other case can be handled in a similar fashion. The limit hypothesis on g allows us to assume that it is a positive function. Moreover, since g' is continuous and nonvanishing it too must be positive.

Let ε be some positive number. Choose M so that

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \varepsilon$$

whenever $x > M$. Since $g'(x)$ is positive we have

$$|f'(x) - Lg'(x)| < \varepsilon g'(x),$$

so that

$$\left| \int_a^b [f'(x) - Lg'(x)]\, dx \right| \leqslant \int_a^b |f'(x) - Lg'(x)|\, dx < \int_a^b \varepsilon g'(x)\, dx$$

whenever $M < a < b$. Therefore, for such a and b,

$$|f(b) - f(a) - L[g(b) - g(a)]| < \varepsilon[g(b) - g(a)]. \qquad (*)$$

Dividing through by the positive number $g(b)$ we obtain

$$\left| \frac{f(b)}{g(b)} - \frac{f(a)}{g(b)} - L\left[1 - \frac{g(a)}{g(b)}\right] \right| < \varepsilon\left[1 - \frac{g(a)}{g(b)}\right] < \varepsilon.$$

It follows easily that

$$\left|\frac{f(b)}{g(b)} - L\right| < \varepsilon + \frac{|f(a)|}{g(b)} + |L|\frac{g(a)}{g(b)}.$$

As b increases, $g(b)$ grows larger and larger without bound. Consequently, both $|f(a)|/g(b)$ and $|L|g(a)/g(b)$ will eventually become (and remain) smaller than ε, implying that

$$\left|\frac{f(b)}{g(b)} - L\right| < 3\varepsilon$$

for all b sufficiently large. This shows that $\lim_{x\to\infty} f(x)/g(x) = L$. □

Since our proof of this version of L'Hôpital's rule makes no use of the assumption $\lim_{x\to\infty} f(x) = \infty$, that condition can be dropped from the hypotheses. A straightforward variation of the preceding proof works when $x \to x_0$; alternatively, that case can be derived from the $x \to \infty$ case by considering $F(x) = f(x_0 + 1/x)$ and $G(x) = g(x_0 + 1/x)$.

This type of proof also applies to the indeterminate form $0/0$. For example, if we assume that L is finite and g' is a positive function (as was the case above), then allowing b to increase without bound in inequality $(*)$ will force $f(b)$ and $g(b)$ towards 0, implying that

$$|-f(a) - L[-g(a)]| \leqslant \varepsilon[-g(a)].$$

Dividing by the (positive) number $-g(a)$ reveals that

$$\left|\frac{f(a)}{g(a)} - L\right| \leqslant \varepsilon$$

whenever $a > M$.

As you can see, the algebra for $0/0$ is a bit simpler making this proof even more suitable for popular consumption.

Acknowledgment. The author wishes to thank the referee for helpful comments on the arrangement of the proofs.

polynomial we mean a line tangent to the graph of the polynomial at more than one point. For example, the x-axis is a multiple tangent of $x^4 - 2x^2 + 1$.)

(a) Prove that $M(n) \leq \binom{n-2}{2}$.

(b)* Prove or disprove: $M(n) = \binom{n-2}{2}$.

(Cf. E 2886 [1981, 349; 1984, 312].

E 3424. *Proposed by Paul Erdös, Hungarian Academy of Science, Budapest.*

(i) Given an integer $r > 1$, prove that there is a positive integer n_r such that for every $n > n_r$ at least one of the binomial coefficients $\binom{n}{k}$, $1 \leq k \leq n - 1$, is divisible by the r-th power of some prime.

(ii) Prove that n_2 can be taken as 23. (The binomial coefficients $\binom{n}{k}$, $1 \leq k \leq n - 1$, are all squarefree when $n = 2, 3, 5, 7, 11, 23$.)

E 3425. *Proposed by Shmuel Rosset and Tamir Shalom, Tel Aviv University, Israel.*

Suppose A and B are n by n matrices over a field k.

(i) Prove that there is an n by n matrix X over k with $AX + XA = B$ if and only if trace $(BC) = 0$ for every n by n matrix C with $AC = -CA$.

(ii) Assume char$(k) = 0$. If $AB = -BA$ and if the matrix equation $AX + XA = B$ has a solution over k, prove that B is nilpotent; for each positive integer n give an example in which $B^n = 0$ but $B^{n-1} \neq 0$.

E 3426. *Proposed by J. Michael Steele, Princeton University, Princeton, NJ.*

Suppose that $\{a_k\}_{k=1}^{\infty}$ is a sequence of non-negative real numbers such that $\sum a_k^2 < \infty$. Show that, for any positive constant c, there exists an increasing sequence of positive integers $\{n_k\}_{k=1}^{\infty}$ such that $n_k = ck^2 + O(k)$ and such that

$$\sum_{k=1}^{\infty} a_{n_k}^2 \sum_{j=1}^{n_k} a_j < \infty.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

The Thinnest Annulus Containing Four Points

E 3296 [1988, 953]. *Proposed by Robert M. Young, Oberlin College, Oberlin, OH.*

Given four arbitrary points in the plane, describe a procedure for constructing the thinnest annulus containing all four points. (An infinite strip is regarded as an extreme case of an annulus.) The thickness of an annulus is the difference between the two radii (or the width of the infinite strip).

Solution by Allen J. Schwenk, Western Michigan University Kalamazoo, MI. We will show below that a minimum thickness annulus must have two of the four

points on each bounding circle (or each side of the strip). Given this, we need only consider the three possible pairings of two points for each boundary. For each pairing, we use the fact that the center of any circle through two points lies on the perpendicular bisector of the segment between them. Thus, we determine the bisectors for the two pairs in a pairing, and their intersection is the center of the corresponding annulus. Finally, we compare the thickness of the three resulting annuli. If the bisectors are parallel, then they intersect at infinity and the annulus is an infinite strip. If the bisectors are identical, then the four points are cocircular or collinear and the minimum thickness is zero; for if the common bisector is taken as the y-axis in a Cartesian coordinate system and if the points have coordinates $(\pm a, 0)$ and $(\pm b, h)$, then (when $h \neq 0$) the points lie on a circle with center at $(0, c)$, where $c = (b^2 + h^2 - a^2)/(2h)$.

Compactness assures that the minimum thickness is attained. To see that each boundary circle must contain (at least) two of the four points, note first that each must contain at least one of the points, or else we can expand or shrink, reducing the thickness, until it does. Now, suppose that only one of the four points, A, lies on the outer (inner) circle of a given minimizing annulus with center O. Then we can translate O toward (away from) A by an appropriate distance ε, simultaneously shrinking (expanding) both circle radii by ε to avoid changing the thickness, so that at most one of the points lies on the other circle, with equality only if the line OA contains that point. If the inner (outer) circle now contains no points, we can expand (shrink) it to reduce the thickness. If it contains one of the four points, say B with OBA collinear, then we can rotate the center around whichever of B or A is between the other and O. Having done this, the outer circle now contains none of the points and can be shrunk. The argument is essentially the same in the case of an infinite strip.

Editorial comment. All solvers noted that the possible locations of the center are at the intersections of two perpendicular bisectors each obtained from two of the four points. Schwenk, Harvey Lipman, and the Western Maryland Group reduced the possibilities to the three described above; the others included the four possibilities in which one point is left out and one of the circles contains the other three points.

When there are more than four points, the argument that each circle of the thinnest bounding annulus must contain at least two of the points holds as before. This yields a quartic algorithm (as a function of the number of points) to find the minimum bounding annulus, but this is not best possible, and it is reasonable to ask for the complexity of finding the thinnest bounding annulus. See T. J. Rivlin, Approximation by Circles, *Computing* 21(1979), 93–104.

Solved also by N. Cohen (Israel), H. Lipman, O. P. Lossers (The Netherlands), T. L. McCoy, A. Nijenhuis, Victor Pambuccian, A. Pedersen (Denmark), D. Secrest, Western Maryland College Problem Group, and the proposer.

Caught in a Pigeonhole

E 3312 [1989, 155]. *Proposed by Robert Blakley and Douglas Hensley, Texas A & M University College Station, TX.*

Given positive integers m, n with $n/2 < m < n$, suppose n tennis matches are to be scheduled on m successive days, with at least one match per day. Determine

the largest positive integer K such that, no matter how the matches are scheduled, for each $k \leqslant K$ there is a consecutive sequence of days on which a total of exactly k matches are scheduled.

Solution material prepared by the editors. Variants and simplifications of this problem appear in almost every textbook that presents the pigeonhole principle. In most versions, the total number of matches is at most n rather than exactly n, and the argument requested suffices to show that a consecutive subsequence totalling k is forced for any $k \leqslant 2m - n - 1$, although a minor refinement of this argument also works for $k = 2m - n$. To wit, if a_j denotes the total number of matches played on the first j days, then the numbers $a_1, \ldots, a_m$ and $a_0 + k$, $\ldots, a_{m-1} + k$ (where $a_0 = 0$) are $2m$ integers in the set $\{1, \ldots, n + k - 1\}$ (since $a_{m-1} < n$), so that $k \leqslant 2m - n$ guarantees a duplication of the form $a_i + k = a_j$, $0 \leqslant i < j \leqslant m$.

In "Some pigeonhole principle results extended", this MONTHLY, 87 (1980) 648–651, Joan P. Hutchinson and Paul B. Trow gave a complete characterization of which values of k are forced to arise as the total number of matches played in some block of consecutive days. They show that k is forced if and only if $r < 2m - n$, where $m \equiv r \pmod{k}$ with $0 \leqslant r < k$. A concise presentation of their proof follows. See also references [1] and [2] below.

THEOREM. *For fixed m, n with $n > m$, let A be the set of positive integers k such that every sequence of m positive integers summing to at most n has a consecutive segment summing to k. Then $k \in A$ if and only if $r < 2m - n$, where $m \equiv r \bmod k$ with $0 \leqslant r < k$.*

Proof. If $r \geqslant 2m - n$, then the sequence alternating $k - 1$ ones and one $k + 1$ until m values are reached has sum $2m - r \leqslant n$ and avoids k. To prove the converse, it suffices to show that $2m - r$ is the minimum sum of all sequences of length m that avoid k. We prove this by induction on m, with $m = 0$ a trivial basis. If $m > 0$ and m is not a multiple of k, then the sum must be at least one more than the sum of the initial segment of length $m - 1$, which also avoids k. Now suppose $m = tk$. If a sequence avoids k and sums to less than $2m$, then the partial sums $a_1, \ldots, a_m$ consist of m numbers from $S = \{1, \ldots, 2m - 1\} - \{k\}$, which has size $2m - 2$. Partition S into $m - 1$ pairs by pairing x and $x + k$ whenever $\lfloor x/k \rfloor$ is even. By the Pigeonhole Principle, a uses two elements from a pair, which implies that the sequence cannot avoid k. □

Since the present problem requires the sum to equal n, its feasible sequences are a subset of those in the Hutchinson-Trow formulation, so any value forced in that formulation is also forced in the current problem. Hutchinson and Trow observe that a number larger than m cannot be forced, as shown by a sequence of all 1's. For the problem at hand, such a sequence is forbidden, since we require the total number of matches to equal n. Dean S. Clark and James T. Lewis consider this precise situation in their paper "Avoiding-sequences with minimum sum," *Discrete Applied Mathematics*, 22(1988/89), 103–108, independently of the earlier work. If $n = m + 1$, then every feasible sequence has a single 2 and m 1's, and every value through $m + 1$ is forced. For any other k satisfying $r \geqslant 2m - n$ with $m \equiv r \bmod k$, $0 \leqslant r < k$, the Hutchinson-Trow sequence that avoids having k as a consecutive sum totals to at most m, but the sequence can be modified to achieve the total n without forcing k. The sequence consists of $k - 1$ ones, then one

$k + 1$, then $k - 1$ ones, etc. The sequence comes in groups of k items totaling $2k$ with r trailing ones. Hence the total is $2(m - r) + r \leqslant n$. We can augment any of the values equaling $k + 1$ to make the sequence total n without losing the property of avoiding k.

Based on the Hutchinson-Trow result, the solution of the problem proposed is

$$K(m, n) = m + 1 \quad \text{if } n = m + 1,$$

$$K(m, n) = \max\{k : m \bmod j < 2m - n \quad \text{for} \quad j = 1, 2, \ldots, k\} \quad \text{otherwise.}$$

(Here $m \bmod j$ means the least non-negative integer congruent to m modulo j.) Hutchinson and Trow note numerous special cases of m, n in which a more "exact" or at least more satisfying solution can be given, but for general m, n no better answer than that above is known. The values through $2m - n$ are always forced, as are those from $m - x$ through m if $x = 2m - n - 1$; this means that whenever $n \leqslant 3m/2$ the answer is m. For $n > 3m/2$ the answer is less than m.

REFERENCES

1. R. L. Hemminger and B. D. McKay, Integer sequences with proscribed differences and bounded growth rate, *Discrete Math.*, 55 (1985) 255–265.
2. W. D. Wei and C. L. Liu, Solution to a problem of J. P. Hutchinson and P. B. Trow, *J. Math. Res. and Exposition*, 3 (1986) 139–140.

Lewis and Clark contributed the reference to their recent solution. Several readers contributed the usual elementary argument that begins the discussion above, and there were several incorrect solutions.

A Greedy Change-Making Machine

E 3323 [1989, 357]. *Proposed by Thane Plambeck, Stanford University, Stanford, CA.*

A change-making machine gives change in chips of three integer denominations, 1 cent, α cents, and β cents, where $1 < \alpha < \beta$. The machine uses the following greedy algorithm: In making M cents change it dispenses chips of value β as long as possible, then chips of value α as long as possible, then chips of value 1 to complete the total of M. We say that a pair $\alpha < \beta$ is frugal if, for every M, the machine's algorithm dispenses the minimum number of chips among all ways of making change for M cents with denominations $1, \alpha, \beta$. (For example, $(2, 5)$ is frugal but $(4, 9)$ is not, because the $(4, 9)$ machine dispenses four chips in change for 12 cents instead of the minimum of three.)

For fixed $n > 2$, suppose the values $\alpha < \beta$ are two distinct integers chosen at random from $\{2, 3, \ldots, n\}$ with all pairs equally likely. Obtain an asymptotic formula (when n is large) for the probability that (α, β) is frugal.

Solution by Richard Stong, Harvard University. Cambridge, MA. Let $F(n)$ be the number of frugal pairs (α, β) with $2 \leqslant \alpha < \beta \leqslant n$. We will show that the desired probability is

$$p_n = \frac{2F(n)}{(n-1)(n-2)} = \frac{8}{3}n^{-1/2} + O(n^{-1}).$$

More specifically, $F(n) = 4k^3/3 - 5k^2/2 + k/6 + 1 - (k^2 - n)(2k - 3)$ for

$k(k-1) \leqslant n \leqslant k^2$ and $F(n) = 4k^3/3 - 5k^2/2 + k/6 + 1 + (n-k^2)(2k-2)$ for $k^2 \leqslant n \leqslant k(k+1)$.

We begin by showing that (α, β) is non-frugal if and only if there is an integer t such that $(t-1)\alpha < \beta \leqslant t(\alpha - 1)$. If there is such a t (the inequality implies $1 < t < \alpha$), then for $t\alpha$ the (α, β) machine dispenses one β and $t\alpha - \beta$ ones (by the first inequality), which exceeds the t copies of α that are optimal (by the second inequality). Conversely, suppose M is a value for which the (α, β) machine is not optimal. It can fail only by using too many β's, which means that the minimum must be achieved by changing s copies of β and i copies of 1 into r copies of α, where $i < \alpha$, $s + i > r$, and $s\beta + i = r\alpha$. Substituting the two inequalities for i into the equality yields $(r-1)\alpha < s\beta < r(\alpha - 1) + s$. Now we divide by s and set $t = \lceil r/s \rceil$. Because $\lfloor (r-1)/s \rfloor = \lceil (r-s)/s \rceil$, we have $(t-1)\alpha < \beta < t(\alpha - 1) + 1$.

We next use this characterization to count, for a given value β, the number of values α such that (α, β) is frugal; this equals $\lceil 2\sqrt{\beta} \rceil - 3$. The characterization implies that a non-frugal (α, β) satisfies $\beta/t + 1 \leqslant \alpha < \beta/(t-1)$ for some integer $t \geqslant 2$. Since we also have $t < \alpha$, we may assume that $2 \leqslant t \leqslant k$ if β is in the interval $(k(k-1), k(k+1)]$. For β in this range, we have $k-2$ intervals that α must avoid for (α, β) to be frugal. Although α cannot be higher than the first interval ending at $\beta - 1$, it can be lower than the last, and this yields $\lfloor (\beta - 1)/k \rfloor - 1$ values of α with $2 \leqslant \alpha < \beta/k + 1$ such that (α, β) is frugal. Between the interval for t and the interval for $t+1$ we have $\beta/t + 1 > \alpha \geqslant \beta/t$, and there is exactly one such integer α, so we have $k-2$ additional values of α such that (α, β) is frugal. Since $\lfloor (\beta - 1)/k \rfloor$ is $k-1$ for $k(k-1) < \beta \leqslant k^2$ and is k for $k^2 < \beta \leqslant k(k+1)$, we obtain a total of $2k-3$ or $2k-2$ values of α in the two ranges. This is summarized by the formula $\lceil 2\sqrt{\beta} \rceil - 3$ for the number of frugal pairs with fixed β, as claimed.

This yields the formula $F(n) = \sum_{b=3}^{n} (\lceil 2\sqrt{b} \rceil - 3)$. Thus $F(n) = 4n^{3/2}/3 + O(n)$, which yields the asymptotic formula claimed for the probability that a pair is frugal (since there are $(n-1)(n-2)/2$ pairs to choose from).

If $k(k-1) \leqslant n \leqslant k^2$, then

$$\begin{aligned} F(n) &= (n - k(k-1))(2k-3) + \sum_{b=3}^{k(k-1)} \left(\lceil 2\sqrt{b} \rceil - 3 \right) \\ &= (n - k(k-1))(2k-3) + \sum_{j=4}^{2k-1} (j-3)\lfloor j/2 \rfloor \\ &= (n - k(k-1))(2k-3) + \sum_{r=2}^{k-1} (2r-3)r + \sum_{r=2}^{k-1} (2r-2)r \\ &= (2k-3)n - \frac{2}{3}k^3 + \frac{1}{2}k^2 + \frac{1}{6}k + 1. \end{aligned}$$

A similar calculation for $k^2 \leqslant n \leqslant k(k+1)$ yields the second formula claimed for $F(n)$.

Editorial comment. Paul Zwier characterized the frugal pairs for fixed α. These consist of all (α, β) with $\beta > \alpha^2$ and the pairs $(\alpha, k\alpha - j)$ such that $k \in \{2, 3, \ldots, \alpha\}$ and $j \in \{0, 1, \ldots, k-1\}$. This yields another way to calculate $F(n)$

and translates directly to Joel Spencer's statement of the condition for frugality: With j, k being the integers defined uniquely by the conditions $\beta = k\alpha - j$ and $0 \leqslant j < \alpha$, the pair (α, β) with $\alpha < \beta$ is frugal if and only if $j < k$. The proposer gave the reference L. Chang and James F. Korsh, "Canonical coin changing and greedy solutions," *J. Assoc. Comp. Mach.*, 23 (1976), 418–422, in which similar problems are discussed.

Solved also by H. Lipman, J. Spencer, P. J. Zwier, and the proposer. Partially solved by M. Bazant (student).

A Deviation From A Binomial Coefficient Identity

E 3327 [1989, 445]. *Proposed by Stanley Rabinowitz, Westford, MA.*

Suppose n is a positive integer greater than 3. Put

$$F_n(x) = \sum_{k=0}^{n} \left| \binom{n}{k} - x^k \right|, \qquad S_n = \min_{x>0} F_n(x).$$

(a) Prove that

$$c2^n < S_n < F_n(1) = 2^n - n - 1$$

for a suitable positive c.
(b) Is it true that $\lim_{n\to\infty} 2^{-n} S_n = 1$?

Solution by Ilan Kozma, Tel Aviv, Israel. (a) By definition, $S_n \leqslant F_n(1)$. To prove the strict inequality $S_n < F_n(1)$ for $n > 3$, choose $x_0 = n^{1/n}$. For $1 \leqslant x < x_0$ and $1 \leqslant k \leqslant n-1$, we have $\binom{n}{k} - x^k \geqslant n - x^n > n - x_0^n \geqslant 0$; hence $F_n(x) = x^n - 1 + \sum_{k=1}^{n-1}\left(\binom{n}{k} - x^k\right)$ for $1 \leqslant x < x_0$. The derivative of this expression is $F_n'(x) = nx^{n-1} - \sum_{k=1}^{n-1} kx^{k-1}$. Since $F_n'(1) = n - \binom{n}{2}$, which is negative for $n > 3$, we conclude that $S_n \leqslant F_n(1+\varepsilon) < F_n(1)$ for sufficiently small positive ε. The existence of c satisfying $S_n > c2^n$ follows from (b).

(b) We shall prove that the answer is yes by using the following lemma.

LEMMA. *If n is sufficiently large and $0 \leqslant k \leqslant m = \lfloor 2n/3 \rfloor$, then*

$$\binom{n}{k} \geqslant 2^k.$$

Proof of Lemma. If $n \geqslant 2$ and $1 \leqslant k \leqslant (n+1)/3$, then

$$\binom{n}{k} \Big/ \binom{n}{k-1} = \frac{n-k+1}{k} \geqslant 2$$

and so the desired inequality follows by induction for $1 \leqslant k \leqslant (n+1)/3$. On the other hand, if $(n+2)/3 \leqslant k \leqslant m$, then $\binom{n}{k} \geqslant \binom{n}{m}$ and so it suffices to prove that $\binom{n}{m} \geqslant 2^m$; but if n is sufficiently large, Stirling's formula gives

$$\binom{n}{m} > \frac{c}{n^{1/2}} \left(\frac{3}{4^{1/3}}\right)^n > 2^{2n/3} \geqslant 2^m$$

for a suitable positive constant c. Thus the lemma is proved.

We return to the proof of the assertion of (b). If $x \geqslant 2$, then

$$F_n(x) \geqslant \left| \sum_{k=0}^{n} x^k - \sum_{k=0}^{n} \binom{n}{k} \right| \geqslant \sum_{k=0}^{n} 2^k - 2^n = 2^n - 1 > F_n(1).$$

Hence we need consider only $0 < x < 2$. By the lemma we have for $0 < x < 2$ and n large

$$\begin{aligned} 2^{-n}F_n(x) &> 2^{-n} \sum_{k=0}^{m} \left| \binom{n}{k} - x^k \right| \\ &\geqslant 2^{-n} \sum_{k=0}^{m} \binom{n}{k} - 2^{-n} \sum_{k=0}^{m} 2^k. \end{aligned} \qquad (*)$$

As n goes to infinity, the second term in (*) approaches zero exponentially, while the Law of Large Numbers implies that the first term approaches 1. Hence $\lim_{n\to\infty} 2^{-n}S_n = 1$.

Editorial comment. Instead of using the Law of Large Numbers in the above proof, we could use the explicit estimate

$$1 - 2^{-n} \sum_{k=0}^{m} \binom{n}{k} = 2^{-n} \sum_{k=m+1}^{n} \binom{n}{k} \leqslant 2^{-n} n \binom{n}{m+1} = O\left(2^{-n} n^{1/2} \left(\frac{3}{4^{1/3}}\right)^n\right)$$

given by Stirling's formula.

Solved also by H. Ki (Korea), H. Lipman, and the editors.

Embedding a Planar Triangulation with Vertices at Specified Points

E 3341 [1989, 642]. *Proposed by P. Gritzmann, University of Siegen, West Germany; B. Mohar, University of Ljubljana, Yugoslavia; J. Pach, New York University and Mathematical Institute of the Hungarian Academy of Sciences; and R. Pollack, New York University.*

Let G be a planar graph obtained from a convex n-gon by triangulating its interior, i.e., by subdividing the interior into triangles using non-crossing diagonals. Let P be an arbitrary set of n points in the plane no three of which lie on a line.

Show that G can be straight-line embedded on P, i.e., show that there exists a bijection f from the vertex set of G to P such that if (a, b) and (c, d) are distinct edges of G, then the open segments $(f(a), f(b))$ and $(f(c), f(d))$ are disjoint.

Solution by the proposers. We establish a stronger statement by induction on n: if v_1, v_2 are any two consecutive vertices of the convex n-gon from which G is obtained, and p_1, p_2 are consecutive vertices of the convex hull of P, then G can be straight-line embedded on P such that $f(v_1) = p_1$ and $f(v_2) = p_2$.

The statement is trivial for $n = 3$; assume $n > 3$. Let $v_1, v_2, \ldots, v_n$ denote the counterclockwise ordering of the vertices of G on the original n-gon, and let v_i be the third vertex on the triangle containing v_1, v_2. We claim there is a point $p \in P - \{p_1, p_2\}$ with the property that (a) no point of P is inside p_1p_2p and (b) there is a line l through p that separates p_1 from p_2, meets P only at p, and has exactly $i - 2$ points of P on the side of l containing p_2. To obtain p, we rotate the

line p_1p_2 about p_2 until we reach a line $l' = p_2p'$ with $p' \in P$ such that exactly $i - 3$ points of P are separated from p_1 by l'. Among the points of $P - \{p_1, p_2\}$ in the closed halfplane determined by l' that contains p_1, let p be the point minimizing the angle p_2p_1p. By this choice, p satisfies (a), and there are at most $i - 2$ points of P on the side of p_1p containing p_2. If we rotate this line about p, then before it becomes parallel to l' it reaches a position l satisfying (b).

Let H_1 and H_2 denote the closed halfplanes determined by l containing p_1 and p_2, respectively. By the induction hypothesis, the subgraphs of G induced by $\{v_2, v_3, \ldots, v_i\}$ and $\{v_i, v_{i+1}, \ldots, v_n, v_1\}$ can be straight-line embedded on $H_2 \cap P$ and $H_1 \cap P$, respectively, so that v_1, v_2, v_i are mapped to p_1, p_2, p. Combining these embeddings yields a straight-line embedding of G with the desired properties.

No other solutions were received.

The Fixed Points of a Permutation

E 3342 [1989, 642]. *Proposed by Irving Adler, North Bennington, Vermont.*

Let $S = \{1, 2, \ldots, n^2\}$ and define a permutation $f: S \to S$ as follows. Take n^2 cards numbered from 1 to n^2 and lay them in a square array, with the i-th row containing cards $(i - 1)n + 1, (i - 1)n + 2, \ldots, in$ in order from left to right. Pick up the cards along rising diagonals, starting with the upper left-hand corner. If card j is the k-th card picked up, put $f(j) = k$. For example, if $n = 3$, then $f(1) = 1$, $f(4) = 2$, $f(2) = 3$, $f(7) = 4$, $f(5) = 5$, $f(3) = 6$, $f(8) = 7$, $f(6) = 8$, $f(9) = 9$.

For each value of n both 1 and n^2 are fixed points of f. If n is odd, then $(n^2 + 1)/2$ is also a fixed point of f. Characterize those n for which f has a fixed point not in $\{1, n^2, (n^2 + 1)/2\}$.

Solution by José Heber Nieto, Universidad del Zulia, Maracaibo, Venezuela. The permutation f has a fixed point not in the set $\{1, n^2, (n^2 + 1)/2\}$ if and only if n has at least two different prime factors. We may assume $n > 1$.

Let the rising diagonals be numbered 1 through $2n - 1$, beginning at the upper left-hand corner and proceeding to the lower right-hand corner. By symmetry a number x in diagonal k with $k \leqslant n$ is fixed if and only if $n^2 + 1 - x$ in diagonal $2n - k$ is fixed. It thus suffices to consider numbers in the first n diagonals. The number of the card originally in diagonal k and row i, where $1 \leqslant i \leqslant k \leqslant n$, is $x = (i - 1)n + k - i + 1$. Further

$$f(x) = \sum_{j=1}^{k-1} j + k - i + 1.$$

Hence $x = f(x)$ if and only if

$$(i - 1)n = k(k - 1)/2 \qquad (*)$$

It follows from $(*)$ that each diagonal contains at most one fixed point. Thus f has a fixed point not in the set $\{1, n, (n^2 + 1)/2\}$ if and only if $(*)$ has a solution (i,k) such that $1 < k < n$ and $1 \leqslant i \leqslant n$. Since k and $k - 1$ are coprime, no such solution exists if n is a prime or prime power.

If n is not a prime or a prime power, then $n = rs$, where r and s are coprime and both greater than one. Further, we may assume that s is odd. Since $2r$ and s are coprime, there are integers α and β such that

$$2r\alpha + s\beta = 1 \qquad (**)$$

Now, if the pair (α, β) satisfies $(**)$, so does the pair $(\alpha + st, \beta - 2rt)$ for any integer t. Thus we may assume that $-r < \beta \leqslant r$. In view of $(**)$ we must have $\beta \neq 0, r$ and thus there are two cases to consider.

Case I, $-r < \beta < 0$, $\alpha > 0$. In this case we take $k = 2r\alpha$, $i = 1 - \alpha\beta$.

Case II, $0 < \beta < r$, $\alpha < 0$. In this case we take $k = s\beta$, $i = 1 - \alpha\beta$.

In both cases $k(k-1)/2 = -\alpha\beta rs = -\alpha\beta n = (i-1)n$. Straightforward calculation shows that $3 \leqslant k \leqslant n - 2$ and $2 \leqslant i \leqslant n/2$. Thus we have a fixed point of f in either case.

Editorial comment. Several other solvers also reduced the problem to finding a solution of $(*)$ with $1 < k < n$ and $1 \leqslant i \leqslant n$. B. M. M. de Weger showed that if m is the number of distinct prime divisors of n, then the number of fixed points of f (including the trivial ones) is $2^{m+1} - 1$ if n is odd and 2^m if n is even.

Solved also by D. Callan, K. David & A. Shilepsky, J. Fukuta (Japan), M. Hartman & D. S. Rubin, R. J. Hendel, R. H. Jeurissen (The Netherlands), H. Lipman, O. P. Lossers (The Netherlands), L. F. Martins (student), W Raffke (West Germany), J. H. Steelman, F. van Dyke, J. T. Ward, B. M. M. de Weger (The Netherlands), J. Zhang (China), and the proposer.

Rationals as Products from a Sequence of Rationals

E 3347 [1989, 735]. *Proposed by Vlastimil Dlab, Carleton University, Ottawa, Canada.*

Given an arbitrary natural number N, show that every rational number between 0 and 1 can be written as a finite product of numbers of the form

$$R_n = \frac{n(n+3)}{(n+1)(n+2)}, \qquad n \geqslant N.$$

Solution by Kerry Behler (student), Shippensburg University, Shippensburg, PA. By induction on n, the product $\prod_{j=0}^{n+1} R_{n+j}$ telescopes to $n/(n+1)$. Given a rational number a/b between 0 and 1, we can write $a/b = Na/Nb = \prod_{i=0}^{Nb-Na-1}(Na+i)/(Na+i+1) = \prod_{i=0}^{Nb-Na-1} \prod_{j=0}^{Na+i+1} R_{Na+i+j}$.

Editorial comment. This solution uses many factors; Roger B. Eggleton showed that $O(N^3)$ factors always suffice, and that $\Omega(N^2)$ factors may be needed. John T. Ward noted that the argument can be generalized to other sequences, such as those of the form $R_k(n) = n(n+3k)/(n+k)(n+2k)$.

Solved also by I. C. Bivens & L. R. King, A. Christos (student), R. B. Eggleton (Brunei), K. Ford (student), Z. Franco, M. Getz, J. C. Jacobson, H. Ki (South Korea), N. F. Lindquist, O. P. Lossers (The Netherlands), R. Martin (West Germany), L. F. Martins, T. McDonald, P. P. Ray, J. H. Steelman, R. Stong, J. T. Ward, K. S. Williams (Canada), Western Maryland College Problems Group, and the proposer.

Coprime Sequences Summing to Unity

E 3359 [1989, 928]. *Proposed by Daniel Ullman, George Washington University, Washington, DC.*

Does there exist a finite increasing sequence $1 < a_1 < a_2 < \cdots < a_n$ of positive integers such that $\sum_{i=1}^{n} 1/a_i = 1$ and such that a_i and a_{i+1} are relatively prime for $i = 1, 2, \ldots, n + 1$?

Composite solution by O. P. Lossers, University of Technology, Eindhoven, The Netherlands, Richard Stong, University of California at Los Angeles, and the proposer. Such sequences exist. Here are four examples:

3, 4, 5, 9, 22, 25, 99, 100	(Lossers)
3, 5, 7, 9, 11, 21, 22, 45, 154	(Lossers)
2, 3, 11, 23, 43, 127, 1771, 1807, 32131, 3263442	(Stong)
2, 3, 14, 33, 35, 88, 115, 273, 296, 403, 609, 943, 1062, 1073, 1206, 1519, 2419, 3283	(Ullman)

No other correct solutions were received.

ADVANCED PROBLEMS

6649. *Proposed by D. E. Knuth, Stanford University, Stanford, CA.*

Let P be a monic polynomial of degree m with complex coefficients and let ω be a primitive m-th root of unity. Let A be the m by m matrix in which the element in the j-th row and kth column is

$$A_{jk} = \int_0^\infty (\omega^k t)^j e^{-P(\omega^k t)}\, dt \qquad (0 \leqslant j, k \leqslant m - 1).$$

Prove that A is nonsingular.

6650. *Proposed by Keith Ball, Trinity College, Cambridge, England and Herman J. Tiersma, University of Technology, Eindhoven, The Netherlands.*

Let S be the set of natural numbers k such that every matrix of zeros and ones containing exactly $2k$ ones must have a submatrix containing exactly k ones. (In the language of graph theory S is the set of natural numbers k such that every bipartite graph with $2k$ edges has an induced subgraph with k edges.)

(a) Show that if p is a prime congruent to 1 modulo 5 such that $2p + 1$ is also prime, then $2p$ and $2p + 1$ are not in S.

(b)* Does S contain infinitely many natural numbers? Are the powers of 2 all in S?

6651. *Proposed by Richard L. Bishop and Lee A. Rubel, University of Illinois at Urbana-Champaign.*

Prove that the differential equation

$$z^4 \frac{d^2W}{dz^2} = W^4$$

has no nonconstant entire solutions, but that, for every $R > 0$, it does have a non-constant solution analytic in $\{z: |z| < R\}$.

SOLUTIONS OF ADVANCED PROBLEMS

Some Elliptic Curves Modulo p

6506 [1985, 740]. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let $p > 3$ be a prime and write $f(n)$ for the Legendre symbol (n/p). Prove that

$$\sum_{m=0}^{p-1} f(m^3 + 2)\{1 - f(4m + 5)\} = 1.$$

Solution by Benedict H. Gross, Harvard University, Cambridge, MA. We consider the elliptic curves defined by the plane equations

$$C: y^2 = x^3 + 2,$$
$$E: y^2 = (x^3 + 2)(4x + 5),$$
$$J: y^2 = x^3 - 120x + 506$$

over the field Q of rational numbers. (A good reference for the needed material on elliptic curves is **[2]**, particularly Theorem 2.3 on page 24 and Exercise 5.4 on page 145.)

In the table in **[1**, p. 129] C is the curve denoted by $1^* - 1$ and J is the curve denoted by $77^* - 1$ (see **[1**, pp. 123–125]). In particular the curves C and J are 3-isogenous over $\mathbb{Q}$ and hence over $\mathbb{Z}/p\mathbb{Z}$ for any prime p greater than 3. (Since the discriminants of C and J are $-2^6 \cdot 3^3$ and $-2^6 \cdot 3^5$ respectively, these are the primes where the two curves have good reduction.) Thus C and J have the same number of points in projective space over $\mathbb{Z}/p\mathbb{Z}$ for any prime greater than 3. Now, C and J each have one point at infinity, namely $(w, x, y) = (0, 0, 1)$, and so they have the same number of finite points over $\mathbb{Z}/p\mathbb{Z}$. If $m^3 + 2$ is a non-zero square in $\mathbb{Z}/p\mathbb{Z}$, there are two points on C over $\mathbb{Z}/p\mathbb{Z}$ with x-coordinate m, while if $m^3 + 2$ is zero in $\mathbb{Z}/p\mathbb{Z}$, there is just one point $(m, 0)$ on C over $\mathbb{Z}/p\mathbb{Z}$ with x-coordinate m. Thus the number of finite points on C over $\mathbb{Z}/p\mathbb{Z}$ is equal to

$$\sum_{m=0}^{p-1} \{1 + f(m^3 + 2)\},$$

and similarly for the number of finite points on J over $\mathbb{Z}/p\mathbb{Z}$. Thus for $p > 3$ we

have

$$\sum_{m=0}^{p-1} \{1 + f(m^3 + 2)\} = \sum_{m=0}^{p-1} \{1 + f(m^3 - 120m + 506)\}. \tag{1}$$

On the other hand, as m runs through $(\mathbb{Z}/p\mathbb{Z}) \setminus \{25/4\}$, it is immediate that $n = (32 - 5m)/(4m - 25)$ runs through $(\mathbb{Z}/p\mathbb{Z}) \setminus \{-5/4\}$. Now if we make the substitution $n = (32 - 5m)/(4m - 25)$, we have

$$(n^3 + 2)(4n + 5) = \frac{9}{(4m - 25)^4}(m^3 - 120m + 506).$$

Thus

$$\begin{aligned}
\sum_{n=0}^{p-1} f(n^3 + 2)f(4n + 5) &= \sum_{n \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{-5/4\}} f(n^3 + 2)f(4n + 5) \\
&= \sum_{m \in (\mathbb{Z}/p\mathbb{Z}) \setminus \{25/4\}} f(m^3 - 120m + 506) \\
&= \sum_{m=0}^{p-1} f(m^3 - 120m + 506) - f(9/64) \\
&= \sum_{m=0}^{p-1} f(m^3 - 120m + 506) - 1. \qquad (2)
\end{aligned}$$

To recapitulate, the curves E and J are isomorphic over $\mathbb{Z}/p\mathbb{Z}$, and the number of finite points on E over $\mathbb{Z}/p\mathbb{Z}$ is one less than the number of finite points on J over $\mathbb{Z}/p\mathbb{Z}$.

Combining (1) and (2) gives the assertion of the problem.

REFERENCES

1. Modular Functions of One Variable IV, Proceedings 1972, edited by B. J. Birch and W. Kuyk, Lecture Notes in Mathematics No. 476 Springer-Verlag, 1975.
2. Joseph H. Silverman, The Arithmetic of Elliptic Curves, Graduate Texts in Mathematics No. 106, Springer-Verlag.

Editorial comment. Daniel Grayson has pointed out to us that the 3-isogenies between $C = \{(p, q): q^2 = p^3 + 2\}$ and $J = \{(z, w): w^2 = z^3 - 120z + 506\}$ can be given explicitly, namely

$$z = p + \frac{24(p + 1)}{(p + 2)^2}, \qquad w = q - \frac{24pq}{(p + 2)^3}$$

and

$$p = \frac{z^3 - 12z^2 + 12z + 152}{9(z - 6)^2}, \qquad q = \frac{w(z^3 - 18z^2 + 132z - 376)}{27(z - 6)^3}.$$

Although these explicit maps apparently do not make it possible to give a proof of (1) avoiding the theory of elliptic curves, they do serve to eliminate the dependence of the above solution on the tables in [**1**].

No solutions were received before the customary deadline. The above solution was prepared by Professor Gross at the request of the editors.

Expressing a Positive Integer As the Sum of As Many Distinct Squares as Possible

6574 [1988, 559]. *Proposed by Moshe Laub, Jerusalem, Israel*

Put $M_r = \sum_{j=1}^{r} j^2$ for $r = 1, 2, \ldots$. If $n \in N$ (the set of positive integers), define $m(n)$ to be the unique positive integer r such that $M_r \leqslant n < M_{r+1}$.

(a) Let S_1 be the set $n \in N$ such that n is expressible as a sum of $m(n)$ distinct squares of positive integers. Prove that S_1 has asymptotic density one but that $N \setminus S_1$ is infinite.

(b) Let S_2 be the set of $n \in N$ such that n^2 is expressible as a sum of $m(n^2)$ distinct squares of positive integers. Prove that S_2 has asymptotic density one. Is $N \setminus S_2$ infinite?

Solution I by L. E. Mattics, University of South Alabama, Mobile, AL. (a) First of all, we note that no number n such that $M_r < n \leqslant M_r + 2r$ is in S_1 and thus $N \setminus S_1$ is infinite. For if $n > M_r$ and n is expressible as a sum of r distinct positive squares, then one of the squares must exceed r^2 and so

$$n \geqslant 1^2 + 2^2 + \cdots + (r-1)^2 + (r+1)^2 = M_r + 2r + 1.$$

In the other direction we shall prove that if

$$M_r + 36r^{3/2} < n < M_{r+1} \qquad (*)$$

and r is sufficiently large, then n is in S_1. Since

$$\sum_{k=1}^{r} 36k^{3/2} < 36r^{5/2} = O\left(M_r^{5/6}\right),$$

this would show that S_1 has asymptotic density one. We begin with a lemma.

LEMMA. *Let s be a fixed positive integer greater than 1. Suppose t_N, u_N, v_N are positive integers depending on the positive integer N. Then the number of solutions of*

$$x_1 + x_2 + \cdots + x_s = N \qquad (**)$$

in distinct positive integers different from t_N, u_N, v_N is

$$\frac{1}{(s-1)!} N^{s-1} + O(N^{s-2}).$$

Proof of the Lemma. The number of solutions of $(**)$ in positive integers with no restrictions is

$$\binom{N-1}{s-1} = \frac{1}{(s-1)!} N^{s-1} + O(N^{s-2}).$$

Now the number of solutions $(**)$ in which two x_r are equal is $O(N^{s-2})$ and the number of solutions in which one of the x_i takes a particular value t_N is also $O(N^{s-2})$ uniformly in t_N. Thus the lemma follows.

Now we turn to the proof that if n satisfies $(*)$, where r is sufficiently large, then $n \in S_1$. Put $n - M_r = 2rp + q$ with $0 \leqslant q < 2r$. Put $P = p - 2, Q = q + 4r$ if $p - q$ is even and $P = p - 1, Q = q + 2r$ if $p - q$ is odd. Then $n - M_r = 2rP + Q$, where $2r \leqslant Q < 6r$ and $P - Q$ is even. Since $r^2 + 2r \geqslant n - M_r \geqslant 36r^{3/2}$, we have $r/2 \geqslant P \geqslant 18r^{1/2} - Q/(2r) > 18r^{1/2} - 3$. Choose a positive integer b

such that $b^2 \leqslant Q < (b+1)^2$. Then

$$Q - (b-1)^2 < (b+1)^2 - (b-1)^2 = 4b < 4(6r)^{1/2},$$
$$Q - (b-1)^2 \geqslant (2b-1) \geqslant 2(Q^{1/2} - 1) - 1 \geqslant 2(2r)^{1/2} - 3,$$

and

$$P - (b-1) \geqslant 18r^{1/2} - 3 - (6r)^{1/2} + 1 > 15r^{1/2}$$

for sufficiently large r. Since $P - Q$ is even, there is an integer $m < r/4$ with

$$2m = P - (b-1) - \{Q - (b-1)^2\},$$

where

$$2m > 15r^{1/2} - 4(6r)^{1/2} > 5r^{1/2} > 2(b-1)$$

if r is sufficiently large. Now let $s = 2$ if $Q - (b-1)^2$ is even and $s = 3$ if $Q - (b-1)^2$ is odd. Recall that $Q - (b-1)^2 > 2(2r)^{1/2} - 3$. Thus, if r is sufficiently large, we may apply the lemma with $2N + s = Q - (b-1)^2$ to determine positive integers $j_1 < j_2 < \cdots < j_s$ such that $j_k \neq b-2$, $j_k \neq m-1$, $j_k \neq m$ for $i = 1, 2, \ldots, s$ and

$$Q - (b-1)^2 = \sum_{k=1}^{s} (2j_k + 1).$$

Clearly $j_s < \{Q - (b-1)^2\}/2 < 2(6r)^{1/2} < r$ for sufficiently large r. Thus we have

$$\begin{aligned}
n &= M_r + 2rP + Q \\
&= M_r + (r+b-1)^2 - r^2 + 2r\{P - (b-1)\} + \{Q - (b-1)^2\} \\
&= M_r + (r+b-1)^2 - r^2 + 2r\{[P - (b-1)] - [Q - (b-1)^2]\} \\
&\quad + (2r+1)\{Q - (b-1)^2\} \\
&= M_r + (r+b-1)^2 - r^2 + 4rm + (2r+1)\sum_{k=1}^{s} (2j_k + 1) \\
&= M_r + (r+b-1)^2 - r^2 + (r+m)^2 - (r-m)^2 \\
&\quad + \sum_{k=1}^{s} \{(r + j_k + 1)^2 - (r - j_k)^2\}.
\end{aligned}$$

Since $0 < j_1 < j_2 < \cdots < j_s$ and $j_k \neq m$ for $k = 1, 2, \ldots, s$, the squares r^2, $(r-m)^2, (r-j_1)^2, (r-j_2)^2, \ldots, (r-j_s)^2$ are distinct. Since $j_k \neq m-1$ and $j_k \neq b-2$ for $k = 1, 2, \ldots, s$, the squares $(r+b-1)^2, (r+m)^2, (r+j_1+1)^2$, $(r+j_2+1)^2, \ldots, (r+j_s+1)^2$ are all distinct. Thus n is equal to the sum obtained from $1^2 + 2^2 + \cdots + r^2$ by deleting $s+2$ of the squares and replacing them by $s+2$ squares greater than r^2. Hence if n satisfies $(*)$ and r is sufficiently large, then $n \in S_1$.

(b) If $r = 48a^2$ for some positive integer a and $n = 192a^3 + 3a$, a simple calculation shows that $n^2 = M_r + r/48$, so that

$$M_r < n^2 < M_r + 2r.$$

Thus, if r is of the above form and n is chosen as indicated, n^2 is not expressible as a sum of r distinct positive squares. Hence $N \setminus S_2$ is infinite.

In the other direction we note that if

$$M_r^{1/2} + 32 < n < M_{r+1}^{1/2} \qquad (***)$$

and r is sufficiently large, then n is in S_2; for if $(***)$ holds, then

$$M_{r+1} > n^2 > M_r + 64M_r^{1/2} > M_r + 64\left(r^3/3\right)^{1/2} > M_r + 36r^{3/2},$$

so that $M_{r+1} > n^2 > M_r + 36r^{3/2}$, $n^2 \in S_1$, and $n \in S_2$. Since $32r = O((M_r^{1/2})^{2/3})$, this shows that S_2 has asymptotic density one.

Solution II by the Editors. We prove the generalization of the assertion of the problem in which squares are replaced by k-th powers and M_r becomes $\sum_{j=1}^{r} j^k$. We require the following theorem which can be readily proved by the Hardy-Littlewood method as presented in [1] or [2].

THEOREM. *Given a positive integer $k > 1$ and a positive number ε, every sufficiently large positive integer N can be expressed as a sum of $2^k + 1$ distinct positive k-th powers each of which is greater than $N/\{(2^k + 1)(1 + \varepsilon)\}$.*

For $k = 2$ this theorem (and more) was proved by E. M. Wright, *Quart. J. Math.* (Oxford) 41 (1933) 37–51 and 228–232.

To obtain the first part of (a) we prove that for large r any n satisfying

$$M_r + 2(2^k + 1)\varepsilon r^k \leqslant n < M_{r+1} \qquad (*)$$

can be expressed as a sum of r distinct positive integer k-th powers. For if n satisfies $(*)$, then

$$n - M_{r-2^k-1} \geqslant \sum_{j=0}^{2^k} (r-j)^k + 2(2^k + 1)\varepsilon r^k$$

$$> (2^k + 1)(1 + \varepsilon)r^k$$

if r is large. By the above theorem $n - M_{r-2^k-1}$ can be expressed as a sum of $2^k + 1$ distinct integer k-th powers each greater than r^k. But then $n = (n - M_{r-2^k-1}) + M_{r-2^k-1}$ can be expressed as a sum of $(2^k + 1) + (r - 2^k - 1) = r$ distinct positive integer k-th powers.

To obtain the second part of (a), note that if $n > M_r$ and n is expressible as a sum of r distinct positive integral k-th powers, then one of the k-th powers must exceed r^k and so

$$n \geqslant 1^k + 2^k + \cdots + (r-1)^k + (r+1)^k = M_r + (r+1)^k - r^k.$$

Thus, none of the integers $M_r + 1, M_r + 2, \ldots, M_r + (r+1)^k - r^k - 1$ is expressible as a sum of r distinct positive integer k-th powers and so $N \setminus S_1$ is infinite; in fact the argument shows that the number of elements of $N \setminus S_1$ not exceeding x is greater than $x^{1-1/(k+1)}$ for large x.

To obtain the first part of (b) note that the differences between consecutive integer k-th powers in the interval $[M_r, M_{r+1})$ are less than $kM_{r+1}^{1-1/k}$. This is small compared to

$$M_{r+1} - M_r = (r+1)^k = \left\{(r+1)^{k+1}\right\}^{1-1/(k+1)} > M_{r+1}^{1-1/(k+1)}.$$

Thus "most" integer k-th powers must be in some interval of the form $(M_r + 2(2^k + 1)\varepsilon r^k, M_{r+1})$ and so are expressible as a sum of r distinct positive integer k-th powers by our solution of (a).

To obtain the second part of (b), we recall (thanks to Jacob Bernoulli) that

$$M_r = \frac{r^{k+1}}{k+1} + \frac{r^k}{2} + \frac{k}{12}r^{k-1} + O(r^{k-2})$$

$$= \frac{r}{k+1}\left(r + \frac{k+1}{2k}\right)^k - \frac{k^2-3}{24k}r^{k-1} + O(r^{k-2}).$$

Thus if $r = (k+1)(2ka)^k$ for some positive integer a, then

$$\frac{r}{k+1}\left(r + \frac{k+1}{2k}\right)^k = \{(k+1)(2ka)^{k+1} + a(k+1)\}^k = u^k$$

is the k-th power of an integer u, and further for large a we have

$$0 < u^k - M_r < \frac{k}{24}r^{k-1} < (r+1)^k - r^k.$$

Thus u^k is not expressible as a sum of r distinct positive integer k-th powers. Since a can be any sufficiently large positive integer, the set $N \setminus S_2$ is infinite.

REFERENCES

1. T. Estermann, On Waring's theorem: A simple proof of a theorem of Hua, *Sci. Rep. Nat. Tsing Hua Univ.*, 5A (1948) 226–239.
2. R. C. Vaughan, The Hardy-Littlewood method, *Cambridge Tracts in Mathematics*, 80 (1981) specifically Chapter 2.

No other solutions were received.

Cyclic Permutations of the q-adic Digits

6601 [1989, 454]. *Proposed by Sajal K. Das, University of North Texas, Denton, TX and the late Patrick O'Hara, University of Central Florida, Orlando.*

Given positive integers $q \geqslant 2$ and $k \geqslant 1$ let π_k be the permutation of the set $\{0, 1, 2, \ldots, q^k - 1\}$ defined by

$$\pi_k(i) = qi - (q^k - 1)\lfloor qi/(q^k - 1)\rfloor \text{ if } i = 0, 1, 2, \ldots, q^k - 2,$$

$$\pi_k(i) = i \text{ if } i = q^k - 1.$$

If π_k is factored into disjoint cycles, determine the cycle lengths which occur and the number of cycles of each length.

Solution by B. M. M. de Weger, Universiteit Twente, Enschede, The Netherlands. Let $0 \leqslant i \leqslant q^k - 2$ have the q-adic expansion

$$i = r_0 + r_1 q + r_2 q^2 + \cdots + r_{k-1}q^{k-1},$$

with $r_0, r_1, \ldots, r_{k-1} \in \{0, 1, \ldots q-1\}$, not all equal to $q-1$. Note that

$$qi = r_{k-1} + r_0 q + r_1 q^2 + \cdots + r_{k-2} q^{k-1} + r_{k-1}(q^k - 1),$$

and by

$$0 \leqslant r_{k-1} + r_0 q + r_1 q^2 + \cdots + r_{k-2} q^{k-1} < (q-1)(1 + q + q^2 + \cdots + q^{k-1}) = q^k - 1$$

it follows that

$$\lfloor qi/(q^k - 1) \rfloor = r_{k-1}.$$

Hence

$$\pi_k(i) = qi - (q^k - 1) r_{k-1} = r_{k-1} + r_0 q + r_1 q^2 + \cdots + r_{k-2} q^{k-1}.$$

Thus the mapping $i \to \pi_k(i)$ can be described as a cyclic permutation of the q-adic digits $(r_0, r_1, \ldots, r_{k-1})$ of i. Note that this is also true for $i = q^k - 1$.

It follows at once that π_k^k is the identity, so that cycle lengths can only be divisors of k. Let $d|k$, and let $c(d)$ denote the number of cycles of length d. If i is in a cycle of length d, then its q-adic digits consist of the block $(r_0, r_1, \ldots, r_{d-1})$ repeated k/d times, and there are just d numbers i in the same cycle. The number of blocks $(r_0, r_1, \ldots, r_{d-1})$ is equal to q^d; they correspond to cycles whose length is a divisor of d. Thus we obtain the formula

$$q^d = \sum_{d'|d} d' c(d').$$

The Möbius Inversion Formula followed by division by d yields for every divisor d of k (including 1 and k itself), the following expression for $c(d)$:

$$c(d) = \frac{1}{d} \sum_{d'|d} \mu(d') q^{d/d'}.$$

Here μ is the well known Möbius μ-function.

Editorial comment. In a similar proof, O. P. Lossers avoided Möbius inversion by a direct application of the inclusion-exclusion principle. Stephen M. Gagola, Jr. pointed out that $c(d)$ is also the number of irreducible polynomials of degree d over $GF(q)$ when q is a prime power. A. A. Jagers asserted that π_k is equivalent to a permutation $\tilde{\pi}_k$ with $\tilde{\pi}_k(s) = s^q$ for all $s \in S$, where S is the set of all complex roots of the polynomial $X^{q^k} - X$. The expression $c(d)$ is also given various interpretations in Dickson's *History of the Theory of Numbers* (vol. I, pp. 84–86). The solution printed above was chosen by an editor fond of q-adic expansions. Shorter solutions were provided by a number of readers.

Solved also by R. J. Chapman (United Kingdom), Stephen M. Gagola, Jr., A. A. Jagers (The Netherlands), Marcin E. Kuczma (Poland), O. P. Lossers (The Netherlands), Yoshio Mimura (Japan), John Henry Steelman, Ondrej Šuch (Czechoslovakia), and the proposers.

Roots on the Unit Circle

6602 [1989, 454]. *Proposed by Allen R. Miller, Naval Research Laboratory, Washington, D.C.*

Suppose $0 \leqslant k_1 \leqslant 1$, $0 \leqslant k_2 \leqslant 1$, and ϕ is real; if $k_1 = k_2 = 1$, suppose further that ϕ is not a multiple of 2π. If

$$\alpha = e^{i\phi}\left(e^{i\phi} - k_1k_2\right),$$
$$\beta = k_1^2 + k_2^2 - 2k_1k_2e^{i\phi},$$
$$\gamma = 2\left(k_1^2 + k_2^2 - k_1k_2\cos\phi - 1\right),$$

prove that the roots of the self-inversive quartic equation

$$\alpha z^4 + \beta z^3 + \gamma z^2 + \bar{\beta} z + \bar{\alpha} = 0$$

have absolute value one.

Solution by David Secrest, Princeton University, Princeton, NJ. Call the polynomial on the left $F(z)$. We shall set $z = e^{i(\theta - \phi)}$ and show that

$$f(\theta) = F(z)/(2z^2)$$

is zero for four values of θ, counting multiplicities. Now

$$f(\theta) = A(\theta) + \overline{A}(\theta),$$

where

$$2A(\theta) = \left(1 - k_1k_2e^{-i\phi}\right)e^{2i\theta} + \left(\left(k_1^2 + k_2^2\right)e^{-i\phi} - 2k_1k_2\right)e^{i\theta} + k_1^2 + k_2^2 - 1 - k_1k_2\cos\phi,$$

so

$$\begin{aligned} f(\theta) &= \cos 2\theta - k_1k_2\cos(2\theta - \phi) + \left(k_1^2 + k_2^2\right)\cos(\theta - \phi) \\ &\quad - 2k_1k_2\cos\theta + k_1^2 + k_2^2 - 1 - k_1k_2\cos\phi \\ &= \cos 2\theta - 1 + \left(k_1^2 + k_2^2\right)(1 + \cos(\theta - \phi)) - 2k_1k_2\cos\theta \\ &\quad - 2k_1k_2\cos(\theta - \phi)\cos\theta \\ &= -2\sin^2\theta + \left(k_1^2 + k_2^2 - 2k_1k_2\cos\theta\right)(1 + \cos(\theta - \phi)). \end{aligned}$$

We claim that it suffices to establish the result for $0 \leqslant \phi \leqslant \pi$. If $\pi < \phi < 2\pi$, substitute $\phi = 2\pi - \phi'$ and $\theta = 2\pi - \theta'$. This gives the same expression with ϕ' and θ' in place of ϕ and θ, but with $0 < \phi' < \pi$. We take the four solutions for θ' and substitute them back to get the four solutions for θ.

Now define θ_0 by $0 \leqslant \theta_0 \leqslant \pi/2$ and $\cos\theta_0 = k_1k_2$. We first consider the case in which $k_1 \neq k_2$ and ϕ is distinct from $0, \pi$, and θ_0. Observe that the following inequalities are valid:

$$f(0) = (k_1 - k_2)^2(1 + \cos\phi) > 0, \tag{1}$$

$$\begin{aligned} f(\theta_0) &< -2\sin^2\theta_0 + 2\left(k_1^2 + k_2^2 - 2k_1k_2\cos\theta_0\right) \\ &= -2\left(1 - k_1^2\right)\left(1 - k_2^2\right) \leqslant 0, \end{aligned} \tag{2}$$

$$f(\pi) = (k_1 + k_2)^2(1 - \cos\phi) > 0, \tag{3}$$

$$f(\pi + \phi) = -2\sin^2(\pi + \phi) < 0. \tag{4}$$

Since $f(\theta)$ is continuous and periodic, there must be four distinct zeros interlaced with 0, θ_0, π, $\pi + \phi$, and 2π. Also, if $\phi = \theta_0$ we still have $f(\theta_0) < 0$ if $k_1 \neq 1$ and $k_2 \neq 1$: the strict and nonstrict inequalities of (2) simply interchange.

To complete the proof, recall that the roots (and hence the absolute values of the roots) of a *monic* polynomial depend continuously on its coefficients. Thus for $\phi \neq 0$ we easily deduce the result by continuity even if $k_1 = 1$, $k_2 = 1$, $k_1 = k_2$, or $\phi = \pi$.

If $\phi = 0$ then

$$F(z) = (z + 1)^2\big((1 - k_1k_2)z^2 + \big(k_1^2 + k_2^2 - 2\big)z + (1 - k_1k_2)\big).$$

Here the roots are $-1, -1, +1, +1$ if $k_1 = k_2$. Otherwise it is easy to show that there is a pair of complex conjugate roots, each of which is the reciprocal of the other. Hence each has absolute value one, and the proof is completed.

Editorial comment. No simple factorization of $F(z)$ was found in general, but Secrest did observe that for $k_1 = 1$, $k_2 = \cos\theta_0$, and $\phi = \theta_0$ we have

$$F(z) = i(\sin\theta_0)(z - 1)^2\big(e^{i\theta_0}z^2 + i(\sin\theta_0)z - e^{-i\theta_0}\big),$$

and that for $k_1 = k_2 = k$ we have

$$F(z) = e^{i\phi}(z - e^{-i\phi})^2G(z),$$

where

$$G(z) = \{(e^{i\phi/2} + k)z + e^{-i\phi/2} + k\}\{(e^{i\phi/2} - k)z + e^{-i\phi/2} - k\}.$$

No other solutions were received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

The Mathematics of Projectiles in Sport. By Neville de Mestre, Australian Mathematical Society Lecture Series 6. Cambridge University Press, 1990. xi + 175 pp.

VICTOR G. SZEBEHELY
Department of Aerospace Engineering and Engineering Mechanics, University of Texas, Austin, TX 78712

Past. I have heard many times that good ideas *might* emerge when the mind is allowed to wander. So I stopped reading this book and suddenly found myself in high school listening to two of my physics teachers. I was 16 when one of my teachers told us to concentrate on the basic principles in mechanics and dynamics and have fun applying these to practical problems. The following year, another teacher forced us to memorize equations by numbering them and the examination went like this. Teacher: "Write equation 16." Student: "I am not sure if it is $v = \sqrt{2gh}$ or $s = (g/2)t^2$." Teacher: "Your second guess is correct. You get a C."

Today I feel sorry for the second teacher and I am glad that he did not succeed in making me hate mechanics. I am grateful for having the first teacher who made me fall in love with dynamics and whom I relate to Professor de Mestre.

Then my mind went to the first year's first semester mechanics classes at the University where only those problems were discussed which could be solved, usually by some specific trick.

If you knew the trick you got an A, if not you were in trouble. I know many professors who emphasize tricky problem-solving instead of the principles.

The second semester offered dynamics and the professor told us that he would not concentrate on solvable problems since they are boring but would apply the basic methods and principles to *real* problems. This was an intellectual challenge which I could not resist and I still jump on "unsolvable and unsolved problems."

Present. At this point my mind stops wandering and I continue reading de Mestre's jewel. He teaches mechanics in a clear, understandable language to some very lucky students whose reaction is to fall in love with dynamics. "How can dynamics be that straightforward and almost too easy?" My mind wanders off again recalling my students whose response to most problems is to run to their computers. So I ask them if their answers are correct? How to verify them? Why not use some approximations which allows them to get a quick (no-computer) answer? And if x is small, why not write $x/2 + 1$ for $\sqrt{x+1}$?

Well, Professor de Mestre tells about approximate solutions of exact equations. He explains the technique of perturbations which, in his competent hands, becomes an undergraduate subject. And he shows graphical integration which helps the reader to understand what is really going on.

When he goes to the concept of mathematical modelling of real problems he addresses engineers who make their living in this way. Which terms shall we keep and which neglect? How will approximate modelling influence the results?

Our limitations regarding analytical solutions versus numerical integration teach us honesty and offer realistic evaluations of our basic problems. Professor de Mestre calls attention to problems not yet treated in the literature, which puts the book on a graduate level, offering subjects for theses or dissertations.

It is often asked, when text books are reviewed, "for whom is the book written"? This is probably the easiest question to answer regarding de Mestre's book: for *everybody*. The undergraduate student will learn how to attack problems in mechanics, his professor will learn how to teach mechanics. The graduate student will understand the technique of approximations and his professor will learn about unsolved problems. Sports enthusiasts will not be able to resist the charts and tables and they might even conclude that it is often easier to compute the trajectory of a football than to kick it with a required speed and direction.

The author states that "The aim of the current book is to present a unified collection of the many problems that can be tackled and of all the mathematical techniques that can be employed." Here he shows his bravery, since "all the mathematical techniques" expand and change every day.

I feel that the aims of this book are to separate the dilettanti from the cognoscenti, to show how to do mechanics well, to explain what mechanics and dynamics are about and, last but not least, how to say it in a beautifully short book satisfying the dictum that less is more. The profession is indebted to Professor de Mestre for his delightful book which treats a special set of dynamical systems of considerable interest.

A Transition to Advanced Mathematics. By Douglas Smith, Maurice Eggen, and Richard St. Andre. Third Edition. Brooks/Cole, Pacific Grove, CA, 1989. xiii + 285 pp.

Foundations for Advanced Mathematics. By Carol Avelsgaard. Scott, Foresman and Company, Glenview, IL, 1990. viii + 325 pp.

Introduction to Advanced Mathematics. By William Barnier and Norman Feldman. Prentice Hall, Englewood Cliffs, NJ, 1990. xv + 320 pp.

JUDITH PALAGALLO
Department of Mathematical Sciences, University of Akron, Akron, OH 44325

Mathematics has been described as a discipline in which early courses give little indication of what will be required in the future and reveal nothing of what is required of a practicing mathematician. During the last few years renewed interest has been shown in "bridging the gap" between the traditional calculus course and more advanced work such as abstract algebra or real analysis. One might ask why this gap exists. The gap has become more prevalent over the years because of the nature of the calculus courses. The theoretical content and rigor of the courses has diminished and the emphasis is placed on computational skills and techniques. Linear algebra has been the course in which the student learns methods of proof

and is introduced to another level of mathematical abstraction. Many times, however, the linear algebra course is also reduced to a methods and skills type of mathematics course. Thus students still have little exposure to the world of mathematics before they enter an analysis or algebra class where they are asked to function like budding mathematicians.

These textbooks make a good attempt at putting together topics needed by students but generally not taught in any other traditional course in an undergraduate curriculum. Each requires the students to have some calculus background. The similarities of the books in content and perspective perhaps indicate that we are precisely aware of what has been lacking in the training of our students. Topics covered by all three textbooks include fundamentals of logic, methods of proof, sets, relations, functions and some counting techniques—material which is indispensable to any mathematician. Beyond these topics the authors diverge into various applications of the principles presented earlier.

The goals of the books are two-fold. First, they desire to provide a firm foundation in the major ideas needed for continued work in mathematics, drawing on the fundamental relationship between logic and mathematics. Second, they intend to guide the students to think and to express themselves mathematically. In building the foundations, the books differ. Barnier and Feldman discuss the propositional calculus in the most sophisticated manner, using quite formal notation. They emphasize symbolic logic before discussing its relationship to methods of proof. The other two books integrate mathematical statements into their examples of logical statements from the beginning, and the presentations are less formal. After the discussions of formal logic, all the books present the standard proof techniques. Good examples of uncomplicated proofs concerning elementary number theory, geometry or beginning calculus are presented. My students prefer the number theory proofs and find any proof involving the concepts of limit or continuity much more difficult. However, it is these proofs that provide the "bridge" we are seeking to build in their mathematical education.

The real test of understanding the methods of proof is whether students can apply the methods to other topics fundamental in their background. As Paul Halmos has stated, "Every mathematician agrees that every mathematician must know some set theory." Our students are no exception, so a detailed discussion of set theory in this course is appropriate. Students are acquainted with sets, but their understanding of set theory is very limited. The next step is to have students practice their newly acquired skills in formal statement-proving by establishing results about sets. Arbitrary families of sets, indexed sets and power sets are new topics and require a deeper level of understanding by the students. Similar remarks could be made about the study of relations and functions. Equivalence relations and inverse images of functions are new topics of some mathematical sophistication. They are also examples of topics which help in the transition from calculus to upper division courses.

I was particularly interested in how each book handled the topic of induction. All students at this level have previously worked with elementary induction problems, but this course should provide an opportunity to relate induction to more appealing properties of the natural numbers. Barnier and Feldman, and Avelsgaard discuss induction among other proof techniques. They use a rather traditional approach, requiring the students to verify that various equations are

true for the natural numbers. Both books do give at least one example of a physical situation which must be modeled by an inductive statement. While these problems are difficult to formulate, they are the ones which give insight into the real nature of the inductive process. Students need more problems presented in such a way. This type of problem requires the student to formulate the inductive statement and verify it with the principle of mathematical induction. Smith, Eggen and St. Andre introduce the topic in a chapter on sets. They take a somewhat different approach by integrating the discussion of induction with the Well Ordering Principle and even the Division Algorithm for the natural numbers. This approach gives the students a look at some real topics in mathematics at a higher level of abstraction. It also demonstrates how a mathematician might discover relationships between topics.

All three books are well written and each contains features worth noting. The Avelsgaard book contains interesting bibliographic references at the end of each chapter, most of which are accessible to the undergraduate student. The books by Smith, Eggen and St. Andre, and Barnier and Feldman both have "proofs to grade" in the exercise sets of the chapters. These are lists of claims with alleged proofs which the students are to evaluate. They expose the students to different styles of writing proofs and develop in the students the ability to read proofs critically. The Barnier and Feldman book has a collection of good supplementary exercises which should not be overlooked. The book by Smith, Eggen and St. Andre ends with three chapters of more advanced mathematics including a discussion of the Axiom of Choice, some group theory and a few important topics from real analysis. Similarly, Barnier and Feldman conclude with four chapters on cardinality, Boolean algebra, the structure of the integers, limits and real numbers. These chapters supply a rich source of examples of the previous material, but they exceed the level of rigor of the earlier chapters. I also question whether there would ever be time in one term to touch on such topics.

Few of the students who take such a course will become professional mathematicians. Nonetheless, the understanding of the logical structure of the language of mathematics and the ability to apply deductive reasoning will be valuable assets to aspiring computer and physical scientists and certainly to those who will eventually teach mathematics. Judging from these books, and others like them, some good attempts are being made to aid students in making the transition from calculus to more advanced mathematics courses.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, C. *Derive: A Mathematical Assistant Program, Third Edition.* Albert Rich, Joan Rich, David Stoutemyer. Soft Warehouse, 1988, vi + 126 pp, (P). *Derive* is a computer algebra system that runs on IBM PC's and compatibles. Major changes between Version 1.22 and Version 1.6 include greater support for multivariable calculus, the addition of a large collection of special functions, improvements to the antidifferentiation algorithm, functions for solving ordinary differential equations, and functions for Fourier series approximations and Laplace transformations. AO

General, T(14-15). *Foundations of Higher Mathematics: Exploration and Proof.* Daniel Fendel, Diane Resek. Addison-Wesley, 1990, xviii + 457 pp, $32.76. [ISBN: 0-201-12587-0] A book whose aim is to bridge the gap between calculus and upper-division courses. Topics covered include logic, set theory, methods of proof, basic number theory, equivalence relations, order, cardinality, limits and continuity, and elementary group theory. JPH

General, S, L. *Innumeracy: Mathematical Illiteracy and Its Consequences.* John Allen Paulos. Penguin Books, 1988, 135 pp, (P). [ISBN: 0-14-012255-9] British paperback reprint of Paulos' 1988 surprise U.S. bestseller (TR, April 1989; Extended Review, January 1990). Innumerable anecdotes reveal appalling public innumeracy, often with serious consequences for personal decisions. LAS

General, P. *Chairing the Mathematical Sciences Department of the 1990s.* National Research Council, 1990, ix + 97 pp, (P). Presentations (with audience comments and questions) from the October 1989 meeting of mathematics department chairs sponsored by the Board on Mathematical Sciences of the National Research Council: dealing with administrations; funding opportunities and strategies; curricular issues; statistics programs. Speakers deal primarily with topics relevant to research universities. LAS

Elementary, T(13: 1). *Fundamentals of Trigonometry, Second Edition.* Charles D. Miller, Margaret L. Lial, David I. Schneider. Scott Foresman, 1990, xiv + 418 pp, $33. [ISBN: 0-673-38961-8] Basic trigonometry text. Key features include an early discussion of circle functions, emphasis on calculator usage throughout, and chapters on complex numbers, exponential functions, and analytic geometry. Provides a set of computer-assisted tutorials and approximately 50 transparencies as supplements. MPR

Mathematics Appreciation, T(13-14). *A Mathematics Sampler: Topics for the Liberal Arts, Second Edition.* William P. Berlinghoff, Kerry E. Grant. Ardsley House, 1988, xiii + 437 pp, $33.95 (P). [ISBN: 0-912675-18-7] A book with genuine mathematical content in each area sampled: number theory, geometry, probability and statistics, computers (Basic programming), infinity (countability), groups, the fourth-dimension. Each chapter closes with the subheading LINK that describes an area of applications. Worth consideration, but a bit unwieldly with 437 pages sized 8 1/2 x 11. AWR

Mathematics Appreciation. *Goop & Gook or 1.f(math) = laughter.* Paul Yearout. Sympun Publishing (2174 N. 250 E., Provo, UT 84604), 1990, 64 pp, $5 (P). A slim collection of mathematical puns, rebuses, and old saws like $\sin x/n = 6$. Possibly useful as a source of humorous filler for departmental newsletters. BC

Education, P*. *Making Mathematics Work for Minorities: Framework for a National Action Plan 1990-2000.* National Research Council, 1990, 24 pp, (P). Highlights of issues and recommendations that emerged from an April 1990 national symposium sponsored by the Mathematical Sciences Education Board. Introduces the agenda of the new Alliance to Involve Minorities in Mathematics (AIMM). LAS

Education, S, P, L. *Assessing Higher Order Thinking in Mathematics.* Ed: Gerald Kulm. AAAS, 1990, 209 pp, $24.95 (P). [ISBN: 0-87168-356-3] Twelve papers analyzing new approaches to assessing school mathematics. Topics include problem solving, computer-based testing, cognitive perspectives, the role of calculators, and quantitative reasoning.

A useful and timely elaboration of issues raised in the NCTM *Standards* which will become increasingly important as the U.S. seeks to measure progress in school mathematics. LAS

Education, S, P, L*. *Constructivist Views on the Teaching and Learning of Mathematics.* Eds: Robert B. Davis, Carolyn A. Maher, Nel Noddings. Journal for Research in Math. Edu., No. 4. NCTM, 1990, ix + 210 pp, $9.50 (P). [ISBN: 0-87353-300-3] Twelve papers delving into the evidence for constructivism—the notion that children construct mathematical knowledge for themselves—and its impact on teaching and learning. One corollary: as a "right answer" is no evidence of correct work, so a correctly written algorithm is no evidence of appropriate understanding. Much emphasis is placed on the importance of developing a "mathematical community" to sustain important emotional factors such as persistence, self-confidence, and trust that are necessary for students to succeed in mathematics. LAS

Education, P, L*. *Changing the Odds: Factors Increasing Access to College.* Sol H. Pelavin, Michael Kane. College Entrance Examination Board, 1990, ix + 85 pp, (P). [ISBN: 0-87447-396-9] Longitudinal analysis of data from 16,000 members of the high school class of 1982 to determine factors that correlate with college attendance and graduation. Key finding: 80% of students who take high school geometry go to college, regardless of ethnic background. No other variable in the study is as effective as geometry in leveling this statistical playing field. Of course, only correlation is documented, not causation. Nonetheless, this report merits close study since it conveys a powerful message. LAS

Education, P*. *Mathematical Challenges for the Middle Grades From the Arithmetic Teacher.* Ed: William D. Jamski. NCTM, 1990, v + 48 pp, $6.25 (P). [ISBN: 0-87353-296-1] Includes 125 problems that appeared either "in form or concept" in the *Arithmetic Teacher* during the last fifteen years. Includes references to the articles that contained these problems and to other relevant articles. JNC

History, C, L**.** *Glory and Failure.* Michael Lindgren. Transl: Craig G. McKay. MIT Pr, 1990, 414 pp, $45. [ISBN: 0-262-12146-8] Social and technical analysis of the first attempts to mechanize the production of numerical tables. History of the difference engine from its conception by Johann Müller in 1784 to the efforts of Charles Babbage (1791-1871) in the years 1823-1833 to the eventual production in 1843 by Georg Scheutz (1785-1873) and his son Edvard Scheutz (1821-1881). Tells why Babbage failed technically and why the Scheutzes failed commercially. Substantial technical descriptions. Superb bibliography of original source material. JDEK

History, S, L*. *The Last Problem.* E.T. Bell. MAA, 1990, 326 pp, $17.50 (P). [ISBN: 0-88385-451-1] A casual, selective, idiosyncratic history of the cultures that gave birth to Fermat's Last Theorem, from Babylonia to seventeenth century France—the age of Fermat. Numerous digressions on remote topics (e.g., Cleopatra's dalliances; Louis XIV taxes) remind us of Bell's reputation for historical fiction. Originally published in 1961, this edition is enveloped by a preface and contemporary notes by editor Dudley. The ratio of entertainment to enlightenment in this delightful volume might set a record for mathematics books. LAS

Logic, T(17-18: 1, 2), P. *Model Theory, Third Edition.* C.C. Chang, H.J. Keisler. Stud. in Logic & Found. of Math., V. 73. North-Holland (US Distr: Elsevier Science), 1990, xvi + 650 pp, $118. [ISBN: 0-444-88054-2] The *Third Edition* of this, the original and still standard introductory text in the field, retains its emphasis on classical first-order model theory, which "remains a prerequisite for all of the newer branches of model theory." Changes include expanded treatment of model completeness; introduction to recursively saturated models; added exercises. (*First Edition,* TR, August-September 1974.) RB

Foundations, P. *Categories, Allegories.* Peter J. Freyd, Andre Scedrov. Math. Lib., V. 39. North-Holland (US Distr: Elsevier Science), 1990, xviii + 296 pp, $30.75 (P); $82. [ISBN: 0-444-70367-5; 0-444-70368-3] Category theory with emphasis on its geometric nature and connections to mathematical logic, via logoi (generalizations of the Heyting algebras which replace Boolean algebras in intuitionistic logic), topoi (categories modeled on the properties of the category of sets), and allegories (which are to the theory of binary relations and sets what categories are to functions and sets). RM

Foundations, T(14), L. *Introduction to Advanced Mathematics.* William Barnier, Norman Feldman. Prentice Hall, 1990, xv + 320 pp. [ISBN: 0-13-477084-6] An appropriate text for "Logic and Proof" courses offered at the sophomore level before real analysis or abstract algebra. The intent is to give students an introduction to the structure of theorems and proofs in a more user-friendly environment. The first half of the book consists mostly of the tools of proving theorems; logic, methods of proof, set theory, etc. The second part consists of topics to which the student may apply these newly-developed skills; cardinality, boolean algebra, the integers, and limits and real numbers. This appears to be a well-constructed and useful text. MPR

Foundations, T(17-18: 1, 2), P. *Abstract and Concrete Categories: The Joy of Cats.* JiříAdámek, Horst Herrlich, George E. Strecker. Wiley, 1990, xii + 482 pp, $54.95. [ISBN: 0-471-60922-6] A high-level compendium of definitions and theorems about the language called category theory intended as both advanced text and reference. JAS

Foundations. *Lecture Notes in Mathematics-1432: Recursion Theory Week.* Eds: K. Ambos-Spies, G.H. Müller, G.E. Sacks. Springer-Verlag, 1990, vi + 393 pp, $37 (P). [ISBN: 0-387-52772-9] Selection of papers from the Second Oberwolfach Conference (1989) on recursion theory, which gives an overview of current work in recursion theory. RM

Graph Theory, S(17-18), P. *Fundamentals of Graph Theory.* Alexander A. Zykov. Transl: L. Boron, C. Christenson, B. Smith. BCS Associates, 1990, vi + 365 pp, $48 (P). [ISBN: 0-914351-04-4] Translation with corrections and supplements of a Russian volume which is a recast version of author's *Theory of Finite Graphs,* the first Soviet graph theory text. Contains descriptive introductions to classical and contemporary topics, developing many to research level. Interesting culturally due to differing language, choice of emphasis, exercises, and many valuable references to Soviet research articles (usually with MR or Zbl reference). JPH

Graph Theory, T*(14-16: 2), S, L*. *Pearls in Graph Theory: A Comprehensive Introduction.* Nora Hartsfield, Gerhard Ringel. Academic Pr, 1990, ix + 246 pp, $29.95. [ISBN: 0-12-328552-6] Innovative, introductory text beginning with elementary concepts covers most traditional topics (e.g., colorings, Eulerian and Hamilton cycles, trees, algorithms), but also has clear exposition of unusual and more advanced topics, like the Oberwolfach Problem, infinite graphs, and magic graphs. Especially good chapters on planar graphs and their colorings, graphs on surfaces, Map Color Theorem, and Heawood's Empire Problem. Develops material to substantial level, including open conjectures. JPH

Graph Theory, T(17-18: 2), P. *Eulerian Graphs and Related Topics, Part 1, Volume 1.* Herbert Fleischner. Annals of Disc. Math., V. 45. North-Holland (US Distr: Elsevier Science), 1990, xiii + 387 pp, $94.75. [ISBN: 0-444-88395-9] Intended for a broad range of readers, including undergraduate students and researchers. Can be read by a mathematician not yet familiar with graph theory, but will also be of interest to researchers since it contains many recent results and partial solutions to some unsolved problems. LC

Graph Theory, S(18), P. *Computational Graph Theory.* Eds: G. Tinhofer, *et al.* Computing Supplementum, V. 7. Springer-Verlag, 1990, 282 pp, $79 (P). [ISBN: 0-387-82177-5] Contains papers edited by G. Tinhofer, E. Mayr, H. Noltemeier, and M. Syslo on computational techniques applied to graph theoretical problems. Includes planar graphs, path problems, coloring problems, orders and graphs, data structures, probabilistic behavior of algorithms, parallel algorithms, VLSI-structures. AD

Combinatorics, T(16). *Introduction to Lattices and Order.* B.A. Davey, H.A. Priestley. Cambridge University Pr, 1990, viii + 248 pp, $18.95 (P); $49.50. [ISBN: 0-521-36766-2; 0-521-36584-8] Covers ordered sets, lattices, Boolean algebras, and some applications. Particular attention paid to distributive lattices. Minimal prerequisites are group theory and discrete mathematics. Plentiful exercises and supplemental material suitable for beginning graduate students. AD

Discrete Mathematics, T(13-14: 2), L. *Applied Discrete Structures for Computer Science, Second Edition.* Alan Doerr, Kenneth Levasseur. Macmillan, 1989, xxiv + 654 pp. [ISBN: 0-02-329934-7] This text acquaints students with the wide array of topics in a standard discrete structures course, and develops mathematical maturity to a level below that of transition to higher mathematics. "The main change [in this edition] is the addition of a large number of exercises at all levels." (*First Edition,* TR, December 1986.) RB

Discrete Mathematics, P. *Mathematical Preliminaries for Computer Networking.* David Claiborne. Wiley, 1990, x + 180 pp, $32.95 (P). [ISBN: 0-471-51062-9] A non-rigorous survey of mathematical techniques and methods used in data communication protocols, coding schemes and network configurations, by a consultant with electrical engineering background, intended for programmers. Assumes college algebra and some calculus. Chapters on probability, information theory, graph theory, queueing theory, finite state machines, and advanced coding techniques. RB

Number Theory, P*. *Number Theory, Volumes I-II.* Eds: K. Győry, G. Halász. North-Holland (US Distr: Elsevier Science), 1990, $179.50 set, [ISBN: 0-444-70476-0]. *Volume I: Elementary and Analytic,* 573 pp; *Volume II: Diophantine and Algebraic,* 495 pp. Proceedings of the János Bolyai Mathematical Society Conference on Number Theory in Budapest from July 20-25, 1987. Contains most, but not all, of the papers presented at the conference. A comprehensive picture of current research in the field. CEC

Number Theory, S*, P, L. *Number Theory.* National Research Council, 1990, vii + 51 pp, (P). Three superb expositions of key ideas of contemporary number theory presented at a public symposium held in conjunction with National Science Week at the National Academy of Sciences in May 1989: H.W. Lenstra, Jr. on number theory applied to primality testing and cryptology; B. Mazur on number theory as a "gadfly"—a constant source of deep problems; and A.M. Odlyzko on relations between quantum chaos and the Riemann hypothesis. Being written for a general audience, these timely surveys can be read by any mathematician. LAS

Number Theory, P. *Lecture Notes in Mathematics-1434: Analytic Number Theory.* Eds: K. Nagasaka, E. Fouvry. Springer-Verlag, 1990, vi + 218 pp, $22 (P). [ISBN: 0-387-52787-7] Seventeen papers from a Tokyo conference in 1988. BC

Linear Algebra. *Matrix Perturbation Theory.* G.W. Stewart, Ji-guang Sun. Comput. Sci. & Scientific Comput. Academic Pr, 1990, xv + 365 pp, $49.95. [ISBN: 0-12-670230-6] Begun as the translation of a book in Chinese by Sun, this new version emerged. Attention is given to the solution of linear systems, least squares problems, and eigenvalue problems. A quote from the preface deserves inclusion: "We debated whether to include numerical examples. Our decision not to was directed by the increasing availability of interactive systems that manipulate matrices. With these systems it is a small matter to play with a perturbation bound, watching

it perform under a variety of circumstances. Besides such lively exercises, a printed example is a thing of lead and stone." AWR

Linear Algebra, T(14: 1). *Matrices and Linear Transformations, Second Edition.* Charles G. Cullen. Dover, 1990, xii + 318 pp, $7.95 (P). [ISBN: 0-486-66328-0] A corrected republication of the 1972 *Second Edition* originally published by Addison-Wesley (TR, August-September 1972). An attractively priced alternative which covers the standard topics. (1966 Addison-Wesley text, TR, March 1967; Extended Review, August-September 1968.) JNC

Linear Algebra, T(14-15: 1). *Linear Algebra with Applications, Third Edition.* Steven J. Leon. Macmillan, 1990, xv + 458 pp. [ISBN: 0-02-369821-7] The most notable change in this edition is a new appendix containing computational exercises to be solved using MATLAB. Also includes a few new theorems, exercises, examples, applications, and figures. (*First Edition,* TR, February 1982; *Second Edition,* TR, June-July 1986.) JNC

Group Theory, T(16-17: 1), L. *Presentations of Groups.* D.L. Johnson. London Math. Soc. Stud. Texts, V. 15. Cambridge University Pr, 1990, 204 pp, $49.50; $16.95 (P). [ISBN: 0-521-37824-9; 0-521-37203-8] A textbook introducing selected topics in combinatorial group theory, representing an option for an advanced undergraduate algebra course. Presumes knowledge of abstract algebra and linear algebra. Much revised from previous (1976 and 1980) editions (TR, May 1976). New chapters on Nielsen's method, relation modules, quotient computation algorithms. Generous assortment of exercises. RB

Algebra, T*(15: 1), S, L*. *Abstract Algebra, Second Edition.* I.N. Herstein. Macmillan, 1990, xix + 293 pp. [ISBN: 0-02-353822-8] The choice of content in this excellent treatment of the subject remains the same in this edition. Errors have been corrected and some minor rewriting has been done to clarify and expand some of the discussion. (*First Edition,* TR, November 1986; Extended Review, October 1987.) CEC

Algebra, P. *General Algebra 1988.* Ed: Rainer Mlitz. North-Holland (US Distr: Elsevier Science), 1990, 265 pp, $84.50. [ISBN: 0-444-88617-6] Proceedings of a conference in memory of Wilfried Nöbauer in Krems, Austria in August 1988. Includes two survey papers, one on the work of Nöbauer, and eighteen previously unpublished papers from the eighty conference lectures. JPH

Calculus, T(13-14: 3).** *Calculus Two: Linear and Nonlinear Functions.* Francis J. Flanigan, Jerry L. Kazdan. Undergrad. Texts in Math. Springer-Verlag, 1990, xvii + 619 pp, $39.95. [ISBN: 0-387-97388-5] *Second Edition*—clarity enhanced through substantial revisions and major additions by users of *First Edition,* Fristedt, Gray, and Frank of the University of Minnesota (TR, December 1971). Retains strong geometric flavor. Linear algebra setting for a study of linear maps. Clusters of helpful remarks throughout. Delightful extras, interesting exercises, and "challenging problems." Thought-encouraging phrasing of non-routine problems is a plus. Somewhat sophisticated material presented in a non-threatening manner. A book that grows on the reader. Prerequisites include single-variable calculus and a bit of mathematical maturity. JDEK

Real Analysis, P. *Hardy-type Inequalities.* B. Opic, A. Kufner. Longman Scientific & Technical (US Distr: Wiley), 1990, 333 pp, $44 (P). [ISBN: 0-582-05198-3] Lecture notes concerning what conditions (on the domain, weight functions, etc.) insure that Hardy-type inequalities hold for certain classes of functions. MLR

Complex Analysis, T(16-18: 1), S*, P, L*. *Complex Analysis: The Geometric Viewpoint.* Steven G. Krantz. Carus Math. Mono., No. 23. MAA, 1990, xv + 212 pp, $23. [ISBN: 0-88385-026-5] An inviting sequel (eminently suitable for independent study) to a first complex analysis course that introduces tools from Riemannian geometry, notably the Poincaré metric, to shed light on the Schwarz lemma and Picard's and Montel's theorems. Using minimal formalism, the author builds insight based on an effective blend of geometric and analytic methods. Introduces the Carathéodory and Kobayashi metrics and proves Poincaré's theorem that there is no biholomorphic map from the bidisc D^2 in C^2 onto the ball B. LAS

Complex Analysis, T(18: 1, 2), S, P. *Introduction to Holomorphic Functions of Several Variables, Volumes I-III.* Robert C. Gunning. Wadsworth, 1990. *Volume I: Function Theory,* xix + 203 pp, $45.95, [ISBN: 0-534-13308-8]; *Volume II: Local Theory,* xix + 218 pp, $51.95, [ISBN: 0-534-13309-6]; *Volume III: Homological Theory,* xix + 194 pp, $51.95. [ISBN: 0-534-13310-X] Based on, but substantially revised and updated from the author's 1965 text with Hugo Rossi, this is a comprehensive, readable, and inviting introduction, at the graduate level, to multivariable complex analysis. *Volume I* covers aspects of function theory including holomorphic, pluriharmonic, and plurisubharmonic functions; pseudoconvexity and domains of holomorphy; $\overline{\partial}$ problems and Dolbeault cohomology. *Volume II* covers local theory of analytic functions and analytic varieties. *Volume III* treats methods for semilocal and global analysis: sheaf cohomology, Cech cohomology, Stein varieties, holomorphic mappings, etc. No exercises. PZ

Complex Analysis, P. *Fifteen Papers in Complex Analysis.* L.A. Aizenberg, *et al.* AMS Transl., Ser. 2., V. 146. AMS, 1990, viii + 112 pp, $65. [ISBN: 0-8218-3130-5] Translation of papers from the 1985 Russian volume *Multidimensional Complex Analysis.* LAS

Differential Equations, P. *Qualitative Theory of Differential Equations.* Eds: B. Sz.-Nagy, L. Hatvani. North-Holland (US Distr: Elsevier Science), 1990, 683 pp, $166.75. [ISBN: 0-444-87407-0] Proceedings of the Third Colloquium on the title subject

held at the Bolyai Institute in Szeged, August 22-26, 1988. Papers are devoted to asymptotic properties and stability problems in ordinary and functional differential equations. Includes applications to mechanics, physics, biology, and control. Except for survey articles, papers have been refereed and contain complete proofs. BL

Differential Equations, T(15). *Ordinary Differential Equations: An Introduction to Nonlinear Analysis.* Herbert Amann. Transl: Gerhard Metzen. Stud. in Math., V. 13. Walter de Gruyter, 1990, xiii + 458 pp, DM 148. [ISBN: 3-11-011515-8] Translated from German and described as a book for second-year students, this book is a notch above what American students might expect. The second motivating example in the book begins on page 12: "Let $M \subseteq R^n$ be an m-dimensional C^2-manifold (e.g., the two-sphere S^2 in R^3), and let" Linear algebra is a prerequisite, the foundations of the calculus of variations are examined, and the author explains that he always has in mind in this course the theory of evolution equations. AWR

Partial Differential Equations, P. *Harmonic Analysis and Partial Differential Equations.* Eds: Mario Milman, Tomas Schonbek. Contemp. Math., V. 107. AMS, 1990, xi + 129 pp, $38 (P). [ISBN: 0-8218-5113-6] Proceedings of a conference held April 4-5, 1988 at Florida Atlantic University. Ten papers applying harmonic analysis to a wide variety of problems in partial differential equations. BL

Partial Differential Equations, P. *Lecture Notes in Mathematics-1431: The Navier-Stokes Equations: Theory and Numerical Methods.* Eds: J.G. Heywood, *et al.* Springer-Verlag, 1990, vii + 238 pp, $22 (P). [ISBN: 0-387-52770-2] Proceedings of a conference at Oberwolfach, September 18-24, 1988. Of the 23 papers, five address numerical methods. BL

Numerical Analysis, T(16-17: 1, 2), L. *Numerical Linear Algebra and Optimization, Volume 1.* Philip E. Gill, Walter Murray, Margaret H. Wright. Addison-Wesley, 1991, xvii + 426 pp, $46.25. [ISBN: 0-201-12649-4] A unified introduction to modern numerical techniques for solving linear systems and linear least squares problems, for updating matrix factorizations, and for linear and nonlinear programming. AO

Numerical Analysis, S.** *Schaum's Solved Problem Series: 2000 Solved Problems in Numerical Analysis.* Francis Scheid. McGraw-Hill, 1990, v + 698 pp, $19.95 (P). [ISBN: 0-07-055233-9] 2000 solved problems in numerical analysis listed under different topics, with a brief summary in front of each chapter. Introduces a large variety of numerical approximation methods, from elementary to advanced problems, including solving differential equations, algebraic systems, polynomials, integrals, and series. LS

Numerical Analysis, S. *M-matrices in Numerical Analysis.* Günther Windisch. Texte zur Math., B. 115. BG Teubner Leipzig, 1989, 140 pp, 17,50 M (P). [ISBN: 3-322-00710-3] An easy access to M-matrices theory. Explores the relation between nonsingular M-matrices and the discretization of second order linear elliptic and parabolic problems by finite element, finite difference and the methods of lines. Excludes convergence analysis and nonlinear problems. Reviewers copy was missing several pages. LS

Numerical Analysis, P. *Algorithms for Approximation II.* Eds: J.C. Mason, M.G. Cox. Chapman & Hall, 1990, 514 pp, $99.95. [ISBN: 0-412-34580-3] Proceedings of the Second International Conference on Algorithms for Approximation at the Royal Military College of Science, Shrivenham, UK, July 12-15, 1988. Ten invited and 33 contributed papers concern development of algorithms (approximation by splines; polynomials; piecewise polynomials; complex methods; interpolation; smoothing and constant methods) and various applications. Includes a very useful, substantial catalogue of algorithms for approximation by Eric Grosse. DFA

Numerical Analysis, T(17-18: 1), P, L. *Numerical Continuation Methods: An Introduction.* Eugene L. Allgower, Kurt Georg. Ser. in Computat. Math., V. 13 Springer-Verlag, 1990, xiv + 388 pp, $69. [ISBN: 0-387-12760-7] Compares predictor-corrector methods of solving nonlinear systems with simplicial methods. After a significant introduction to both methods, the author makes the argument that these methods, seemingly disparate, are actually quite similar. SM

Numerical Analysis, T(15-17), S, L*. *Numerical Methods That Work.* Forman S. Acton. MAA, 1990, xx + 549 pp, $16.50 (P). [ISBN: 0-88385-450-3] Reprint of the 1970 original edition (TR, January 1971; Extended Review, January 1974). A commonsense approach to fundamental methods that have not changed much in twenty years (despite advances in computation), with effective examples and sage advice to keep novice analysts away from numerical quicksand. First half surveys standard problems and methods (evaluation, roots, interpolation, quadrature, differential equations, eigenvalues, Fourier series); second half revisits topics with greater subtlety and increased dimension. LAS

Operator Theory, P. *Mathematical Analysis and Numerical Methods for Science and Technology, Volume 3: Spectral Theory and Applications.* Robert Dautray, Jacques-Louis Lions. Transl: John C. Amson. Springer-Verlag, 1990, x + 515 pp, $98. [ISBN: 0-387-50208-4] Topics include spectral decomposition of self-adjoint operators and examples in electromagnetism and quantum physics. (*Volume 2: Functional and Variational Methods,* TR, June-July 1989.) MLR

Functional Analysis, P. *Nonstandard Methods in Fixed Point Theory.* Asuman G. Aksoy, Mohamed A. Khamsi. Universitext. Springer-Verlag, 1990, ix + 139 pp, $29 (P). [ISBN: 0-387-97364-8] Self-contained, unified treatment of recent work on applications of Banach space ultraproducts to the fixed point theory of non-expansive mappings (natural extensions of contraction mappings to subsets of Ba-

nach spaces). RM

Functional Analysis, T(17-18: 1), P. *Analyse Harmonique.* M. Samuelides, L. Touzillier. Cepadues-Editions, 1990, 224 pp, 180 F. [ISBN: 2-85428-240-X] A concise introduction beginning with classical Fourier analysis. Covers convolutions, Fourier transforms, harmonic synthesis, distributions, and Laplace transforms. Each brief chapter sketches both key mathematical results and common applications. LAS

Functional Analysis, P. *Lecture Notes in Mathematics-1433: Topics in Nevanlinna Theory.* Serge Lang, William Cherry. Springer-Verlag, 1990, 174 pp, $18 (P). [ISBN: 0-387-52785-0] Nevanlinna theory in the classical case of meromorphic functions and generalizations to equidimensional holomorphic maps $f : C^n \to X$, where X is a compact manifold. Possible supplement for a graduate course in complex analysis. MLR

Functional Analysis, T(16-17), P, L. *Theory of Distributions: A Non-Technical Introduction.* Ian Richards, Heekyung Youn. Cambridge University Pr, 1990, ix + 147 pp, $39.50. [ISBN: 0-521-37149-X] Intended for the "non-specialist" interested in distributions (generalized functions, e.g., the Dirac delta function). Good explanations, nice style. MLR

Analysis, P, L. *Scientific Computation on Mathematical Problems and Conjectures.* Richard S. Varga. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 60. SIAM, 1990, vi + 122 pp, $21.50 (P). [ISBN: 0-89871-257-2] Looks at the interplay between analysis and high-precision calculations in six independent examples: the Bernstein conjecture in approximation theory, the "1/9" conjecture for semi-discrete approximations of parabolic equations, the Riemann Hypothesis, zeros of the partial sums of e^z, real vs. complex rational approximations, and generalizations of Jensen's inequality. BC

Analysis, P. *Tables of Fourier Transforms and Fourier Transforms of Distributions.* Fritz Oberhettinger. Springer-Verlag, 1990, viii + 259 pp, $35 (P). [ISBN: 0-387-50630-6] "New, revised, and enlarged version of the previously published book by the author entitled *Tabellen zur Fourier Transformation* (Springer-Verlag, 1957)." Impressive collection of functions tabulated. BH

Differential Geometry, P. *Lecture Notes in Mathematics-1424: Twistor Theory for Riemannian Symmetric Spaces.* Francis E. Burstall, John H. Rawnsley. Springer-Verlag, 1990, 112 pp, $14.70 (P). [ISBN: 0-387-52602-1] Studies the relationship between flag manifolds and Riemannian symmetric spaces with inner involution. Also studies the structure of the zero set of the Nijenhuis tensor on $J(N)$. Applies results of these studies to harmonic maps of S^2 into a symmetric space N. Offers uniform proof of results of Calabi, Eells-Wood, and Bryant concerning harmonic 2-spheres in S^{2n}, CP^n, and HP^1. Completely characterizes stable harmonic 2-spheres into an irreducible Riemannian symmetric space. Extends results of Uhlenbeck on harmonic 2-spaces in $U(n)$ to other Lie groups. OJ

Differential Geometry, P. *Harmonic Maps and Minimal Immersions Through Representation Theory.* Gabor Toth. Perspectives in Math., V. 12. Academic Pr, 1990, xii + 154 pp, $24.95. [ISBN: 0-12-696075-5] An advanced monograph on the general problem of embedding minimal surfaces in Euclidean space. The present setting concerns "harmonic" embeddings, i.e, embeddings *via* maps whose components are harmonic functions. The author studies properties, especially boundary properties, of the parameter spaces of such maps, known as moduli spaces. PZ

Geometry, P. *Lecture Notes in Mathematics-1436: Enumerative Geometry.* Ed: S. Xambó-Descamps. Springer-Verlag, 1990, v + 303 pp, $32 (P). [ISBN: 0-387-52811-3] Ten papers (plus one by J.M. Miret and S. Xambó Descamps) from twenty-nine lectures presented at a conference in Sitges, Spain, June 1987. JPH

Geometry, T*(15-16: 1), P, L. *Basic Concepts of Geometry.* Walter Prenowitz, Meyer Jordan. Ardsley House (320 Central Park West, New York, NY 10025), 1989, xix + 350 pp, $34.95 (P). [ISBN: 0-912675-48-9] A paperback republication of a classical introduction to modern geometry; begins with incidence geometry and adds postulates to study other geometries including projective, affine, Euclidean, and non-Euclidean. JNC

Geometry, T(14-16: 1), S, P. *Measure, Topology, and Fractal Geometry.* Gerald A. Edgar. Undergrad. Texts in Math. Springer-Verlag, 1990, xiii + 230 pp, $29.95. [ISBN: 0-387-97272-2] Text for a rigorous pure mathematics course on fractal geometry. Fractal examples, metric topology, topological dimension, self-similarity, measure theory, Hausdorff dimension. Prerequisites: experience with mathematical proofs; basic abstract set theory; solid background in calculus. Exercises integrated throughout the chapters. Dozens of illustrations and color plates. RB

Algebraic Topology, P. *Advances in Homotopy Theory.* Eds: S.M. Salamon, B. Steer, W.A. Sutherland. London Math. Soc. Lect. Note Ser., V. 139. Cambridge University Pr, 1989, xiii + 177 pp, $27.95 (P). [ISBN: 0-521-37907-5] Fifteen papers from the Fifth Oxford Topology Symposium held in 1988 at Cortona, Italy to celebrate the sixtieth birthday of Ioan James. JAS

Algebraic Topology, P. *Lecture Notes in Mathematics-1423: Stable Homotopy Groups of Spheres: A Computer-Assisted Approach.* Stanley O. Kochman. Springer-Verlag, 1990, viii + 330 pp, $31.20 (P). [ISBN: 0-387-52468-1] A development of a theoretical basis for efficient computation of the stable homotopy groups of spheres (herein carried out for the first 64 stable stems). Includes a discussion of algorithms used in the Fortran programs. Program listings are available from the author. The computational techniques using spectral sequences may be of more general interest. JAS

Differential Topology, T(18), S, P. *Variational*

Principles of Topology: Multidimensional Minimal Surface Theory. A.T. Fomenko. MIA, V. 42. Kluwer Academic, 1990, xviii + 374 pp, $133. [ISBN: 0-7923-0230-3] Offers new approach to the multidimensional Plateau problem. Uses the concepts of stratified surfaces and their stratified volumes to reformulate and solve it in the context of spectral Bordism theory. Requires a strong foundation in algebraic topology, differential geometry, and the calculus of variations. OJ

Topology, T*(17-18: 1, 2), P. *General Topology, Revised and Completed Edition.* Ryszard Engelking. Heldermann Verlag, 1989, viii + 529 pp, DM 148. [ISBN: 3-88538-006-4] Some updating reflecting current research and additions to the bibliography as well as a few corrections characterize this new version of the 1977 classic text and reference. JAS

Topology, T(15-16), L. *Introduction to Topology, Third Edition.* Bert Mendelson. Dover, 1990, ix + 206 pp, $5.95 (P). [ISBN: 0-486-66352-3] Paperback reprint of the late author's elementary, but thorough, introduction to axiomatic point set topology (essentially the same as the Allyn and Bacon *Second Edition*). Includes also an introduction to homotopy theory, the fundamental group, and surface topology. (1968 Allyn and Bacon *Second Edition,* TR, January 1969; 1975 *Third Edition,* TR, June-July 1975.) JPH

Topology, T(18: 2). *Topics in General Topology.* Kiiti Morita, Jin-iti Nagata. Math. Lib., V. 41. North-Holland (US Distr: Elsevier Science), 1989, xi + 747 pp, $197.25. [ISBN: 0-444-70455-8] A collection of fifteen monographs by various Japanese topologists on recent developments in general topology designed as a text or reference work for advanced students seeking to get to the research frontier. A good common index adds to its value as a reference. JAS

Optimization, T(17-18: 2), P, L. *Optimization and Nonsmooth Analysis.* F.H. Clarke. Classics in Appl. Math., V. 5. SIAM, 1990, xii + 308 pp, $28.50 (P). [ISBN: 0-89871-256-4] A reprint of the 1983 Wiley text (TR, March 1984). Part of a series of "out-of-print" books published by SIAM. Corections have been incorporated into this edition. No exercises. SM

Dynamical Systems, P. *Chaos in dynamischen Systemen.* Willi-Hans Steeb, Albrecht Kunick. Bibliographisches Institut, 1989, 239 pp, (P). [ISBN: 3-411-14152-2] A general introduction, in German, to theory and methods of chaotic dynamical systems. The dynamical systems considered arise in various contexts: difference equations, differential equations, Hamiltonian systems, cellular automata, etc. A number of applications are also considered, mainly to physics, mechanics, and electronics. Contains many diagrams and computer programs. PZ

Control Theory, P. *Lecture Notes in Control and Information Sciences-142: Algebraic Theory of Linear Feedback Systems with Full and Decentralized Compensators.* A.N. Gündeş, C.A. Desoer. Springer-Verlag, 1990, v + 176 pp, $27.10 (P). [ISBN: 0-387-52476-2] Algebraic approach (ring theoretic factorizations) to the study of linear, time-invariant, multi-input/output feedback control systems (full and decentralized, discrete and continuous time systems). Focus on closed loop stability, parametrization of stabilizing compensators. RM

Control Theory, P. *Control of Quantum-Mechanical Processes and Systems.* A.G. Butkovskiy, Yu. I. Samoilenko. Math. & Its Applic., V. 56. Kluwer Academic, 1990, xiii + 232 pp, $99. [ISBN: 0-7923-0689-9] From the Preface: "Purposeful control of the quantum state of matter has long been the focus of investigation for many researchers working in various branches of physics [such as laser technology] This book sums up the results of modern research in the theory of automatic control, which can be used to solve many theoretical problems of control systems defined by the equations of quantum physics." BC

Control Theory, T(16-17: 1, 2). *Mathematical Control Theory: Deterministic Finite Dimensional Systems.* Eduardo D. Sontag. Texts in Appl. Math., V. 6. Springer-Verlag, 1990, xiii + 396 pp, $39. [ISBN: 0-387-97366-4] Accessible treatment of basics of control and systems theory: linear systems, optimal control, dynamic feedback, observers, reachability, controlability, parametrization and stability, deterministic Kalman filtering. Minimal prerequisites, with appendices which cover needed material from linear algebra, normed spaces, ordinary differential equations. RM

Computer Literacy, S, P, L? *Computer Viruses and Anti-Virus Warfare.* Jan Hruska. Ellis Horwood (US Distr: Wiley), 1990, 128 pp, (P). [ISBN: 0-13-171067-2] A handbook which presents operational methods and internal structures of viruses on MS-DOS personal computers, and discusses countermeasures, anti-virus software, etc. Does not treat non-DOS machines or threats other than viruses (e.g., worms, Trojan horses). C language source for two anti-virus programs included; bibliography, anti-virus software vendors. RB

Elementary Computer Science, T(13-14: 1), L. *Great Ideas in Computer Science: A Gentle Introduction.* Alan W. Biermann. MIT Pr, 1990, xx + 446 pp, $27.95 (P). [ISBN: 0-262-52148-2] An introduction to the discipline of computer science that emphasizes ideas and concepts. Focuses on the discoveries and technological breakthroughs that have made computers an important part of our daily life. Covers programming, hardware, and software. AO

Elementary Computer Science, T(13-14), L. *A First Course in Modula-2.* Lowell A. Carmony, Robert L. Holliday. Computer Science Pr, 1990, xii + 419 pp, $29.95 (P). [ISBN: 0-7176-8229-4] A suitable text for a first computer science course. The emphasis is on problem-solving and top-down, structured design of algorithms using the Modula-2 programming language. The tone of the text is informal at first, allowing students to achieve intuition about formal programming procedures. Chapter 8 is dedicated entirely to a discussion of disciplined programming techniques, e.g., top-down designs, pseudo-

codes, commenting, and debugging. Many exercises and examples. MPR

Programming, T(15-17: 1), C, P. *Object-Oriented Programming with Simula.* Bjørn Kirkerud. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xiii + 515 pp, (P). [ISBN: 0-201-17574-6] The aim is object-oriented programming. So why Simula? Because the author likes it and makes a case that Simula is an especially good vehicle for learning object-oriented programming. Sample programs are available by "anonymous `ftp`" from the author's internet host `ifi.uio.no`. JAS

Programming, S, P, L.** *The Annotated C++ Reference Manual.* Margaret A. Ellis, Bjarne Stroustrup. Addison-Wesley, 1990, x + 447 pp, $37.75. [ISBN: 0-201-51459-1] A complete reference for the C++ programming language. It contains the reference manual recently adopted as the basis for ANSI standardization of the language together with annotations and commentary about the design and implementation of the language. Examples are given that explore the differences between C and C++ and that explain language subtleties. AO

Programming, S(15-16), P. *Partial Computation and the Construction of Language Processors.* Frank G. Pagan. Prentice-Hall, 1991, x + 166 pp. [ISBN: 0-13-651415-4] Partial computation is a (hitherto theoretical) programming strategy involving programs writing programs: to solve a problem, Program A uses some of the data to write Program B, which can solve the problem given the remaining data. This monograph introduces a methodology based on this idea, and demonstrates it through the case study of compiler/interpreter construction. RB

Programming, T(13-14: 1), L. *Interactive FORTRAN 77: A Hands-On Approach, Second Edition.* I.D. Chivers, Jane Sleighthome. Ser. in Comput. & Applic. Ellis Horwood (US Distr: Simon & Schuster), 1990, 224 pp. [ISBN: 0-13-466764-6] A new edition of the 1984 text. A new feature is the addition of an appendix listing all the intrinsic functions of Fortran 77. Covers standard material. Many provocative exercises. SM

Programming, S(14-17), P, L. *C++: A Guide for C Programmers.* Sharam Hekmatpour. Prentice-Hall, 1990, x + 264 pp, (P). [ISBN: 0-13-109471-8] An introduction to those features of the C++ language which extend C language, for undergraduates, graduates, and professionals. The added facilities of C++ are described in the first half of the book; the second half presents four case studies as examples of application. Several programming exercises with answers. RB

Programming, S, P. *UNIX System V, Release 4: ANSI C Transition Guide.* AT&T. Prentice-Hall, 1990, iii + 49 pp, (P). [ISBN: 0-13-933698-2] Official System V documentation: eleven technical articles on how to write new and upgrade existing C code to comply with the ANSI C language specification. Originally written by David Prosser of Bell Labs for internal use. RB

Programming, T(15-17: 1), P. *The C++ Workbook.* Richard S. Wiener, Lewis J. Pinson. Addison-Wesley, 1990, xviii + 349 pp, (P). [ISBN: 0-201-50930-X] Intended as a sequel to any standard programming text for C or C++, this book aims to develop both facility and style for the serious user of C++. JAS

Programming, S(13-16). *The C++ Answer Book.* Tony L. Hansen. Addison-Wesley, 1990, ix + 578 pp, (P). [ISBN: 0-201-11497-6] A heavily annotated set of solutions to the exercises in Stroustrup's basic language manual *The C++ Programming Language.* This volume provides a significant body of example code which has the blessings of AT&T people. JAS

Programming, S(14), L. *LISP Style & Design.* Molly M. Miller, Eric Benson. Digital Pr, 1990, x + 214 pp, $26.95 (P). [ISBN: 1-55558-044-0] A primer on design and style of Common Lisp programs, not Lisp itself; assumes reader knows Lisp. For the authors, style comprises efficient algorithms, good organization, appropriate abstractions, well-constructed function definitions, useful commentary, effective debugging. Uses a single large programming example throughout to illustrate the design process. Very readable. DFA

Programming, T(15-17). *Programming in C++.* Stephen C. Dewhurst, Kathy T. Stark. Prentice-Hall, 1989, vi + 233 pp, (P). [ISBN: 0-13-723156-3] More and less than a programming manual. This book explains the ideas behind C++ and provides examples of programming techniques using the language. Not a language manual, although it discusses the features of the language in detail. Aimed at experimental programmers. JAS

Programming, S, L. *Object-Oriented Software.* Ann L. Winblad, Samuel D. Edwards, David R. King. Addison-Wesley, 1990, xviii + 291 pp, $29.25 (P). [ISBN: 0-201-50736-6] An overview for interested persons, no matter what their programming experience—none, some, much. Object orientation and its benefits; object-oriented languages, databases, user-interfaces; topics in object-oriented applications—analysis, design, programming, maintenance, project management, development tools, and environments; examples of applications. Nontechnical. Useful list of references and glossary. DFA

Languages, T(16-17: 1), P, L. *Functional Programming: Practice and Theory.* Bruce J. MacLennan. Addison-Wesley, 1990, xii + 596 pp, $35.50. [ISBN: 0-201-13744-5] An introduction to functional programming methodology covering both practical and theoretical issues. No specific language is used in the book; examples are presented using a pseudocode based on standard mathematical notation. AO

Languages, P, L*. *Applications of Object-Oriented Programming.* Eds: Lewis J. Pinson, Richard S. Wiener. Addison-Wesley, 1990, x + 212 pp, $37.75. [ISBN: 0-201-50369-7] Six case studies that describe the use of object-oriented programming lan-

guages in projects that have been designed and implemented. The focus of each case study is on the object-oriented design, its implementation, and the lessons used. The languages used in the projects include C++, Smalltalk, Objective-C, Actor, and Object Pascal. AO

Languages, T(14-15: 1), P. *Functional Programming with Hope.* Roger Bailey. Ser. in Computers & Their Applic. Ellis Horwood (US Distr: Wiley), 1990, xii + 301 pp. [ISBN: 0-13-338237-0] An introductory text on functional programming using the Hope programming language. Hope combines the semantics of a functional programming language with strong typing, user-definable data structures, and features that support modular program design. AO

Languages, T*(15-17: 1), L. *Programming Languages: An Interpreter-Based Approach.* Samuel N. Kamin. Addison-Wesley, 1990, xi + 640 pp. [ISBN: 0-201-06824-9] A textbook for a programming language course that focuses on non-imperative languages; specifically, Lisp, APL, Scheme, Sasl, Clu, Smalltalk, and Prolog. An interpreter for a simplified subset of each language is included. AO

Languages, S(15-17), L. *Programming with Generators, An Introduction.* Alfs Berztiss. Ser. in Comput. & Their Applic. Ellis Horwood (US Distr: Wiley), 1990, 194 pp. [ISBN: 0-13-739087-4] Generators are similar to procedures, but some local state is preserved between calls. Programming with generators allows one to focus on streams of values rather than on individual values. This book provides examples of the use of generators and develops programming validation and verification techniques appropriate for use with generators. AO

Languages, P, L. *The REXX Language: A Practical Approach to Programming, Second Edition.* Michael Cowlishaw. Prentice-Hall, 1990, xii + 203 pp, (P). [ISBN: 0-13-780651-5] The author developed REXX beginning in 1979 to be a general purpose procedural language that is elegant and easy to deal with. REXX has an interpretive style; it is useful for general purpose computing and as a cross-operating system command language interpreter (most implementations are on IBM systems). This manual officially defines REXX 4.0, summarizes philosophy behind the language. RB

Algorithms, T*(14-15: 1), L. *Algorithms in C.* Robert Sedgewick. Addison-Wesley, 1990, xii + 657 pp. [ISBN: 0-201-51425-7] This is a C language version of the author's popular book *Algorithms.* Presents algorithms for sorting, searching, string-processing as well as for problems in graph theory, geometry, and numerical mathematics. AO

Algorithms, T(15-17: 1, 2), L*.** *Introduction to Algorithms.* Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest. Electrical Eng. & Comput. Sci. Ser. MIT Pr, 1990, xvii + 1028 pp, $49.95. [ISBN: 0-262-03141-8] A comprehensive introduction to the study of computer algorithms. Algorithms are described using a pseudocode and analyzed in terms of run-time efficiency. Each chapter presents an algorithm, a design technique, an application area, or a related topic. Useful as a reference as well as a textbook. AO

Algorithms, S(16-18), P. *Solving Problems On Concurrent Processors, Volume II: Software for Concurrent Processors.* Ian G. Angus, *et al.* Prentice-Hall, 1990, x + 805 pp. [ISBN: 0-13-829714-2] Software codes for scientific algorithms and applications described in *Volume I.* Includes discussion of concurrent systems currently available, benchmark codes, and tutorial introduction to CrOS III (crystalline—i.e., synchronous—communication system for concurrent computers), currently available for Intel, NCUBE hypercubes, and the BBN Butterfly. RM

Computer Systems, T(15-17), S, P, L.** *Inside Smalltalk, Volume I.* Wilf R. LaLonde, John R. Pugh. Prentice-Hall, 1990, xvi + 512 pp. [ISBN: 0-13-468414-1] The first volume of a two-volume in-depth introduction to using Smalltalk-80, an object-oriented programming environment. This volume covers fundamental concepts, the Smalltalk environment, the basic classes, and graphics classes. Assumes the reader has experience programming in a procedural language and has access to a Smalltalk system. AO

Computer Systems, T(14-16: 1), P, L. *UNIX Network Programming.* W. Richard Stevens. Prentice-Hall, 1990, xi + 772 pp. [ISBN: 0-13-949876-1] A hands-on text about Unix network software development, illustrated with 15,000 lines of documented source code. Presumes C reading knowledge, computing experience. Discussions of: Unix processes; interprocess communication mechanisms; overviews of networking and popular protocols; BSD sockets and System V Transport Layer Interface; about a dozen complete examples amplified by discussions of issues. Exercises after most chapters. RB

Computer Systems, P. *OSF/Motif, Revision 1.0.* Open Software Foundation. Prentice-Hall, 1990. *OSF/Motif User's Guide, Revision 1.0,* xii + 66 pp, (P), [ISBN: 0-13-640509-6]; *OSF/Motif Style Guide, Revision 1.0,* xi + 156 pp, (P), [ISBN: 0-13-640491-X]; *OSF/Motif Programmer's Guide, Revision 1.0,* xxiv + 726 pp, (P), [ISBN:0-13-640525-8]; *OSF/Motif Programmer's Reference, Revision 1.0,* ix + 941 pp, (P). [ISBN: 0-13-640517-7] Motif is a set of tools for window management and applications program development in a windowing environment built on top of the X-Window system. Motif is an attempt by the Open Software Foundation to provide a high-level standard for applications running in a windowing environment. The *User's Guide* is a quick tutorial to using the Motif user environment. The *Style Manual* is a specification manual for standardizing the user interface. The other two volumes provided guidance, examples, and function references for the developer of programs intended to run in this environment. JAS

Computer Systems, P. *METAFONTware.* Donald E. Knuth, Tomas G. Rokicki, Arthur L. Samuel.

TEXUsers Group (PO Box 9506, Providence, RI 02940), 1990, $20 (P). A collection of four technical reports describing the most recent versions of the programs GFType, GFtoPK, GFtoDVI, and MFT. The versions described here are those that correspond to Version 1.8 of METAFONT which provided support for 8-bit input and extended ligatures. AO

Computer Systems, P. *X Window System: The Complete Reference to Xlib, X Protocol, ICCCM, XLFD, Second Edition, X Version 11, Release 4.* Robert W. Scheifler, James Gettys. Digital Pr, 1990, xxiv + 851 pp, $49.95 (P). [ISBN: 1-55558-050-5] Complete reference for low levels of the X11 window system Release 4 ("X"), written by participants in the development of X. Contents: Xlib C language library interface (bulk of the text); the X communication protocol; communication conventions between applications (clients), font description specifications. RB

Computer Systems, P, L. *The X Window System in a Nutshell for Version 11 Release 4 of the X Window System: Desktop Quick Reference.* Eds: Daniel Gilly, Tim O'Reilly. O'Reilly & Associates, 1990, xi + 367 pp, $24.95 (P). [ISBN: 0-937175-24-2] A quick reference handbook for users of the X Window System (Version 11, Releases 3 and 4). Primarily of interest to experienced users who are writing programs using Xlib and Xt. AO

Computer Systems, P, L. *XView Programming Manual for Version 11 of the X Window System, Volume Seven.* Dan Heller. O'Reilly & Associates, 1990, xxviii + 557 pp, $30 (P). [ISBN: 0-937175-38-2] XView (X Window-System Visual/Integrated Environment for Workstations) is a user-interface toolkit that supports interactive, graphics-based applications running under the X Window System. This book provides a basic introduction to developing applications using this toolkit and also serves as a reference manual. AO

Computer Systems, S(14-15), C, P, L. *Hypertext and Hypermedia.* Jakob Nielsen. Academic Pr, 1990, xii + 263 pp, $29.95. [ISBN: 0-12-518410-7] Hypertext is software which allows one to access information in a nonsequential manner. At certain points the user has several options as to what to read (access) next. Hypertext is beginning to make an impact in areas such as on-line documentation, computer-aided learning, and other areas with large amounts of interconnected information. This book provides an overview of its uses, implementation, and history. MPR

Computer Systems, P. *X and Motif Quick Reference Guide.* Randi J. Rost. Digital Pr, 1990, vii + 369 pp, $24.95 (P). [ISBN: 1-55558-052-1] Summary reference of critical information in three primary references on the upper and lower layers of the X window system (Release 4) and OSF/Motif (Version 1.1) programming. Not a replacement for an introductory guide or the comprehensive manuals. Function headers; data structure definitions; various tables. RB

Computer Systems, P. *OSF/Motif Reference Guide.* Douglas A. Young. Prentice-Hall, 1990, 155 pp, (P). [ISBN: 0-13-642786-3] Summary reference manual for the Intrinsics layer and the OSF/Motif widget collection for the X windows system, by the author of widely-used introductions to X windows. Not a replacement for introductory manuals or comprehensive references. RB

Computer Systems, P. *X Window System Toolkit: The Complete Programmer's Guide and Specification, X Version 11, Release 4.* Paul J. Asente, Ralph R. Swick. Digital Pr, 1990, xxxv + 967 pp, $44.95 (P). [ISBN: 1-55558-051-3] Comprehensive guide for upper layers (Intrinsics, widgets) of the X11 window system Release 4 ("X"), by the lead designers of those layers; companion volume to *X Window System: The Complete Reference to Xlib, X Protocol, ICCCM, XLFD.* Assumes C language and overview of X architecture. Application programming with the toolkit; writing widgets (bulk of the text); specifications for functions, data structures. RB

Computer Systems, P. *The X Window System: Programming and Applications with Xt, OSF/Motif Edition.* Douglas A. Young. Prentice-Hall, 1990, x + 533 pp, (P). [ISBN: 0-13-497074-8] A tutorial introduction to writing programs in an X Window System environment using Xlib and the Xt Intrinsics library. An updated version of the author's 1989 book that uses the Motif release 1.0 widget set rather than the Xw widget set. AO

Computer Systems, P, L. *UNIX for FORTRAN Programmers.* Mike Loukides. O'Reilly & Associates, 1990, xviii + 244 pp, $24.95 (P). [ISBN: 0-937175-51-X] An excellent book describing the features of the Unix Fortran compiler **f77** and related software. The reader is assumed to know the Fortran language. Discussed here are the Unix editor **vi**, the debugging programs **dbx** and **adb**, the compilation tool **make**, and the large scale source code librarian **rcs**. Some minimal shell programming for the Bourne Shell. Environmental variables for the C-shell are also briefly mentioned. This book is extremely well written. SM

Computer Systems, P, L. *X Window Systems: Programming and Applications with Xt.* Douglas A. Young. Prentice-Hall, 1989, x + 468 pp, (P). [ISBN: 0-13-972167-3] A tutorial introduction to writing programs in an X Window System environment using Xlib and the Xt Intrinsics library. Based on the Xt Intrinsics and the Xw widget set from Releases 2 and 3 of Version 11 of the X Window System. AO

Computer Systems, P, L. *X Toolkit Intrinsics Reference Manual for Version 11 of the X Window System, Second Edition, Volume Five.* Ed: Tim O'Reilly. O'Reilly & Associates, 1990, xii + 764 pp, (P). [ISBN: 0-937175-57-9] A reference manual for the Xt Intrinsics library consisting of the reference pages for the library together with a permuted index, a quick reference guide, and other useful information. AO

Computer Systems, P, L. *X Toolkit Intrinsics*

Programming Manual for Version 11 of the X Window System, Second Edition, Volume Four. Adrian Nye, Tim O'Reilly. O'Reilly & Associates, 1990, xxxiii + 543 pp, (P). [ISBN: 0-937175-56-0] The Xt Intrinsics library consists of routines for building and using *widgets* (pre-built user-interface components) for the X Window System. This book is a tutorial introduction to building applications using existing widgets and writing new widgets. It is based on the Xt Intrinsics and the Xw widget set from Release 4 of Version 11 of the X Window System. AO

Computer Graphics, C. *PC Graphics with GKS.* Peter R. Bono, *et al.* Transl: Cynthia Skripak. Prentice-Hall, 1990, xi + 321 pp, (P). [ISBN: 0-13-654435-5] Introduces programmers to computer graphics using GKS. Provides basic concepts of the graphics system along with examples of GKS functions. Illustrates function definitions in two variations: GKS standard and Fortran 77. Classifies all GKS functions at the end. LS

Computer Graphics, L.** *Graphics Gems.* Ed: Andrew S. Glassner. Academic Pr, 1990, xxix + 833 pp, $49.95. [ISBN: 0-12-286165-5] A collection of algorithms, programs, and mathematical techniques for the computer graphics programmer. It contains brief tutorials and clever ideas from the folklore of computer graphics that deal with nuts-and-bolts issues of programming and implementation. An appendix contains C language implementations of many of the algorithms described. AO

Computer Graphics, P. *X Window System User's Guide for X11 R3 and R4 of the X Window System.* Valerie Quercia, Tim O'Reilly. X Window System Ser., V. 3. O'Reilly & Associates, 1990, xxvi + 723 pp, $30 (P). [ISBN: 0-937175-14-5] An introduction and reference for first-time and experienced users of the X11 windows system ("X"), releases 3 and 4; part of a popular series of X guides. Contents: using X, including the "twm" window manager; customizing X; Unix-style reference pages; appendices on system management, fonts, etc. RB

Theory of Computation, P. *Lecture Notes in Computer Science-425: Algebraic Logic and Universal Algebra in Computer Science.* Eds: C.H. Bergman, R.D. Maddux, D.L. Pigozzi. Springer-Verlag, 1990, xi + 292 pp, $27.60 (P). [ISBN: 0-387-97288-9] Proceedings of a conference in Ames, Iowa (June 1988) in which mathematicians and computer scientists, who had been independently applying similar algebraic methods to their different problem types, sought to identify common ground. Sixteen papers representing a wide range of topics. RB

Theory of Computation, P. *Lecture Notes in Computer Science-440: CSL '89.* Eds: E. Börger, H. Kleine Büning, M.M. Richter. Springer-Verlag, 1990, vi + 437 pp, $35 (P). [ISBN: 0-387-52753-2] Proceedings of the Third Annual Workshop on Computer Science Logic held at the University of Kaiserslautern, October 1989. Twenty-eight papers on a mixture of topics in the theory of logic in computer science. RB

Artificial Intelligence, S(16-18), P. *Minimalist Mobile Robotics: A Colony-style Architecture for an Artificial Creature.* Jonathan H. Connell. Perspect. in Artif. Intell., V. 5. Academic Pr, 1990, xvii + 175 pp, $29.95. [ISBN: 0-12-185230-X] Describes "Herbert," a robot with grasping arm that wanders around an unstructured office environment collecting empty soda cans. Design philosophy uses the world as its own best model, with distributed, independent multi-agent "behaviors," coordinated by a priority-based arbitration scheme, with no centralized control, no complex internal representations, no internal communciation between agents, yet yielding seemingly purposeful behavior. RM

Computer Science, P. *Annual Review of Computer Science, Volume 4, 1989-1990.* Ed: Joseph F. Traub, *et al.* Annual Reviews, 1990, viii + 540 pp, $45. [ISBN: 0-8243-3204-0] This series presents high-quality critical review articles, intending to help define and set standards in the discipline; publication is being suspended for lack of willing authors. Articles on topics including supercomputers, distributed file systems, complexity classes, superscalar architectures, several artificial intelligence fields, architecture, software prototyping. RB

Computer Science, S(15-16), P. *Software Excellence.* Larry Shiller. Prentice-Hall, 1990, xv + 239 pp. [ISBN: 0-13-829748-7] This book defines excellent software as software that meets its design objectives, is manageable, is changeable and long-living, and is learnable by someone not part of the original design team. Develops the idea of excellent software and describes how it can be achieved. AO

Computer Science, P. *Advances in Computers, Volume 30.* Ed: Marshall C. Yovits. Academic Pr, 1990, x + 331 pp, $65. [ISBN: 0-12-012130-1] Survey articles on specialized parallel architectures for textual databases, database design, software reliability, cryptography, Soviet computing. RB

Applications, P. *Conference Proceedings.* TEX Users Group (PO Box 9506, Providence, RI 02940), 1988, vi + 262 pp, (P). Proceedings of the annual conference held at Montreal, August 22-24, 1988. The articles emphasize commercial and production environments as well as solutions for problems for dealing with non-English languages including Hebrew. JAS

Applications (Economics), P, L. *The Collected Economics Articles of Harold Hotelling.* Ed: Adrian C. Darnell. Springer-Verlag, 1990, ix + 178 pp, $49.80. [ISBN: 0-387-97011-8] Harold Hotelling, a statistician with very little formal training in economics, made contributions to economics which, though not voluminous, were pioneering in mathematical economics. Here they all are, together with a 28-page introduction to the life and economic thought of Hotelling. WE

Applications (Engineering), T(14: 2), P. *Advanced Engineering Mathematics, Second Edition.* A.C. Bajpai, L.R. Mustoe, D. Walker. Wiley, 1990, ix + 502 pp, $34.95 (P). [ISBN: 0-471-92595-0]

Aimed at engineering undergraduates. Good review of materials for those seeking professional degrees. Mix of analytical, numerical, and computer-based techniques. Light on "proofs." Numerical techniques approached informally. Plentiful supply of worked examples and problems for solution, with answers to most. Changes from *First Edition* (TR, February 1978) include more on use of calculator and microcomputer. Chapter on integral transforms extended. Discrete Fourier Transform introduced. JDEK

Applications (Fluid Dynamics), P. *Topological Fluid Mechanics.* Eds: H.K. Moffatt, A. Tsinober. Cambridge University Pr, 1990, xviii + 805 pp, $115. [ISBN: 0-521-38145-2] Seventy-six papers from a 1989 conference on topological aspects of fluid mechanics and magnetohydrodynamics. BC

Applications (Fluid Dynamics), T(17-18), P. *Nonlinear Waves, Solitons and Chaos.* E. Infeld, G. Rowlands. Cambridge University Pr, 1990, xi + 423 pp, $85; $29.95 (P). [ISBN: 0-521-32111-5; 0-521-37937-7] Self-contained, graduate-level text concentrating on nonlinear wave phenomena in fluids and plasmas. Starts with linear waves and instabilities. Treats surface waves, weakly and fully nonlinear waves, coherent behavior (like solitons) and noncoherent phenomena such as chaos and turbulence. Exercises following each chapter. Good for researchers entering the field. BL

Applications (Fluid Dynamics), T(15-16: 1), L. *A Mathematical Introduction to Fluid Mechanics, Second Edition.* A.J. Chorin, J.E. Marsden. Texts in Appl. Math., V. 4. Springer-Verlag, 1990, 168 pp, $29. [ISBN: 0-387-97300-1] Section headings say it best. In order, they are Euler's Equations, Rotation and Vorticity, Navier-Stokes Equations, Potential Flow, Boundary Layers, Vortex Sheets, Characteristics, Shocks, The Riemann Problem, Combustion Waves. Based on a course in fluid mechanics at Berkeley. This edition incorporates some updates and revisions. Some exercises added; informal; intentionally references are limited and not exhaustive. JDEK

Applications (Management), S(14-16), P. *Software Project Management For Small to Medium Sized Projects.* John J. Rakos. Prentice-Hall, 1990, xviii + 302 pp. [ISBN: 0-13-826173-3] A text intended primarily for project managers, usable at technical or business schools as well as within companies, generally non-technical in nature, describing a standard approach in which phasing and documentation are simplified to fit smaller projects. Three parts: phases for building a project; practical management methods; managing people. RB

Applications (Physics), T(16-17), S, P, L. *Inflation and Quantum Cosmology.* A.D. Linde. Ed: Robert Brandenberger. Academic Pr, 1990, xii + 199 pp, $29.95. [ISBN: 0-12-450145-1] An authoritative discursive survey organized into three chapters: particle physics and inflationary cosmology provides a brief overview of the inflationary universe scenario and a discussion of the problems with the big bang model; generation of density perturbations discusses in detail the mechanism by which energy density fluctuations can lead to galaxy formation; and quantum cosmology and the stochastic approach to inflation which covers recent developments. It ends with a series of amateurish cartoons, the humor of which is lost on the reviewer. MU

Applications (Physics), S(18), P. *Renormalized Quantum Field Theory.* O.I. Zavialov. Math. & Its Applic., V. 21. Kluwer Academic, 1990, xii + 524 pp, $149. [ISBN: 90-277-2758-9] Extensively revised and updated from the 1979 original *Renormalized Feynman Diagrams,* this new epic contains previously unpublished results. MU

Applications (Physics), S(16-18), P, L. *Spinors and Calibrations.* F. Reese Harvey. Perspectives in Math., V. 9. Academic Pr, 1990, xiii + 323 pp, $39.95. [ISBN: 0-12-329650-1] Primarily a collection of examples: the (simple) Lie groups, the spin groups for general signature, normed algebras for general signature, the exceptional groups G_2 and F_4, the orbit structure of the simpler representations of these groups, and the special Lagrangian and associative calibrations. MU

Applications (Physics), P. *Integrable Systems in Quantum Field Theory and Statistical Mechanics.* Eds: M. Jimbo, T. Miwa, A. Tsuchiya. Adv. Stud. in Pure Math., V. 19. Academic Pr, 1989, 682 pp, $89.50. [ISBN: 0-12-385342-7]

Applications (Physics), P. *Mathematical Analysis and Numerical Methods for Science and Technology, V. 1: Physical Origins and Classical Methods.* Robert Dautray, Jacques-Louis Lions. Transl: Ian N. Sneddon. Springer-Verlag, 1990, xvii + 695 pp, $89.90. [ISBN: 0-387-50207-6]

Applications (Social Science), P. *Chaos and Socio-Spatial Dynamics.* Dimitrios S. Dendrinos, Michael Sonis. Appl. Math. Sci., V. 86. Springer-Verlag, 1990, xvii + 184 pp, $39.80. [ISBN: 0-387-97283-8] Introduces an iterative process which the authors claim has wide applicability to the social and biological sciences, particularly to population dynamics. The discussion is intended to be qualitative, and true to their word, the authors provide few proofs. Readers should prepare for an expositional style more typical of social scientists than of mathematicians. JO

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; AD: Amy Davidow, Macalester; BH: Bruce Hanson, St. Olaf; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; BL: Brian Loe, Carleton; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; LS: Li Sheng, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.



Wadsworth & Brooks/Cole Advanced Books & Software

Ethnomathematics: A Multicultural View of Mathematical Ideas
Marcia Ascher

In this fascinating and ground-breaking new book, Marcia Ascher introduces the mathematical ideas of people in traditional, or "small scale," cultures, such as American Indian tribes, Incas, and various cultures in the South Pacific and Africa. As Ascher explores mathematical ideas involving numbers, logic, spatial configuration, and the organization of these into systems and structures, readers will gain both a broader understanding of mathematics and an appreciation for the ideas of other peoples. 1991. Approximately 224 pages. Casebound. ISBN: 0-534-14880-8. **$36.95.**

Complex Variables, Second Edition
Stephen D. Fisher

This undergraduate text for mathematics or engineering majors offers a direct route to the most important topics in the theory and applications of complex variables. Hundreds of solved examples, exercises, and applications help readers appreciate complex variables as both a practical tool and a mathematical structure of enormous beauty and elegance. 1990. 426 pages. Casebound. ISBN: 0-534-13260-X.

APL with a Mathematical Accent
Clifford A. Reiter and William R. Jones

Providing a concise introduction to APL's powerful mathematical, statistical, and graphical capabilities, this book helps readers master computational skills applicable to linear algebra, statistics, probability, numerical analysis, operations research, abstract algebra, number theory, and mathematical modeling. 1990. 225 pages. Paperbound. ISBN: 0-534-12864-5.

Harmonic Analysis
Henry Helson

This reprint of the 1983 edition is appropriate for graduate students who are familiar with the Lebesgue integral, complex function theory, and Banach spaces. 1991. 192 pages. Paperbound. ISBN: 0-534-15570-7.

A Course in Ring Theory
Donald S. Passman

This textbook for a graduate course in ring theory offers a module theoretic stroll through a mixture of commutative and noncommutative ring theory, with emphasis on the latter. Projective and injective modules supply the framework for a coherent choice of material. 1991. 320 pages. Casebound. ISBN: 0-534-13776-8.

Statistics for Business and Economics
Debra Olson Oltman and James R. Lackritz

Offering a classic, non-calculus-based treatment of descriptive and inferential statistical methods and their application to business problems,this thorough new text helps students develop competence in selecting and applying decision-making procedures they'll need as managers and executives. To emphasize not just problem solving but what needs to be done and said after the answer is obtained, the authors provide more detailed explanations and more interpretation of results than is offered in most textbooks. The book includes many real examples, problems, and case studies, demonstrating how statistics is used in a wide variety of business-related fields. 1991. Approximately 1000 pages. Casebound. ISBN: 0-534-14430-6. *Accompanied by:* Instructor's Solutions Manual. Student Solutions Manual. Data Disk. EXP-TEST.

Probability and Statistics for Engineering and the Sciences, Third Edition
Jay L. Devore

This new edition of Devore's best-selling text continues to use real data in both exercise sets and examples and adds a new chapter on Quality Control. Devore focuses on the probability models and statistical methods students are most likely to encounter and use in their careers in engineering and the natural sciences. 1991. Approximately 704 pages. Casebound. ISBN: 0-534-14352-0.

Theory of Point Estimation
E.L. Lehmann

A reprint of the 1983 edition. 1991. 506 pages. Casebound. ISBN: 0-534-15978-8.

Testing Statistical Hypotheses, Second Edition
E. L. Lehmann

A reprint of the 1986 edition. 1991. 603 pages. Casebound. ISBN: 0-534-15984-2.

Statistics with Stata
Lawrence C. Hamilton

This book/diskette package guides users through Student Stata, a full-featured graphics-oriented statistical package for IBM PCs and compatibles. **Statistics with Stata** can be used at a variety of levels, from introductory to advanced. The book's focus is on how to do the work of data analysis, using Student Stata as a power tool. Hamilton illustrates data analysis with examples and includes interpretive discussion as needed. 1990. 176-page paperbound book plus two 5 1/4" disks. ISBN: 0-534-12850-5.

Wadsworth & Brooks/Cole Advanced Books & Software

State-of-the-art software

New for 1991

MathWriter™ 2.0: The Scientific Word Processor for the Macintosh®
J.Robert Cooke and E. Ted Sobel

MathWriter is a complete "what you see is what you get" scientific word processing program that allows users to enter mathematical expressions as text, not graphics, so both text and mathematics can be edited in the same document. In addition to a spelling checker, thesaurus, and automatic hyphenation, the program also automatically sizes and centers mathematical symbols, italicizes variables, and formats tables and matrices. **MathWriter** shows on-screen renumbering of both equations and references to equations and can show both old and revised versions of the same document.
1991. **Professional version: $395.**
Reduced-feature **Educational Version: $99.95.**

New for 1991

EXP2TeX: The EXP to TEX Conversion Utility, Version 2.0
Tom Morris

This exciting new program allows you to combine the speed and versatility of EXP with the typesetting capabilities of TEX! Write with the truest "what you see is what you get" scientific word processor for the IBM PC, then use *EXP2TeX* to convert your file to a Plain TEX source file. If you use the EXP macros included with this program, most files will not require any editing in TEX.
1991. **$195.**

New for 1991

The EXP Font Development Kit, Version 2.0
Donald C. Mullen, Jr. and Simon L. Smith

This Microsoft® Windows™ 3.0-based program allows you to create new symbols and import soft fonts into EXP, Version 2.0. The package contains the source files for all standard EXP fonts in 8-, 10-, and 12-point sizes.
1991. **$150.**

EXP: The Scientific Word Processor, Version 2.0 (for IBM PCs and compatibles)
Simon L. Smith

Speed, ease of use, and an exceptional ability to handle mathematical fonts and symbols have made EXP one of today's most popular "what you see is what you get" scientific word processors.
1990. **$295.**

To order a personal copy of software or books, use our toll-free number **(800) 354-9706**, or write to us at the address below. To request complimentary copies of books for review, or to receive a copy of our 1991 Advanced Books & Software Catalog, please write:

Wadsworth & Brooks/Cole Advanced Books & Software
MM001
511 Forest Lodge Road
Pacific Grove, CA 93950
(408) 373-0728

When was the last time a computer program helped you think about mathematics?

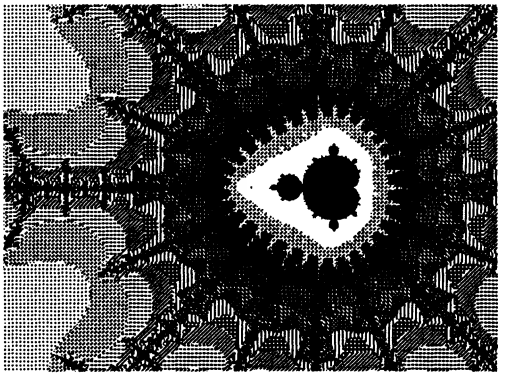
Call or write for our latest software catalog.
For PC or Macintosh.

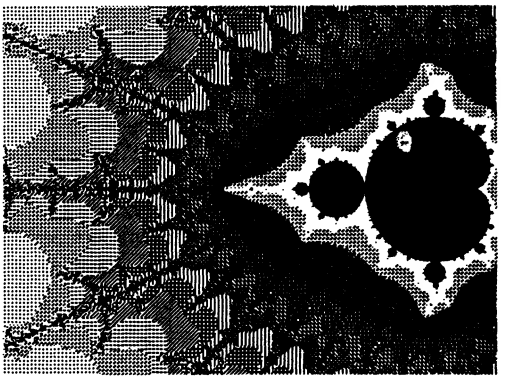
Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

DI-MAN™

FAST Protected Mode Fractal Graphics
Plus...Orbit Diagrams & Frame Animation!
For IBM 386 *w.387 coprocessor,4 MB* **or486 $49.95**

10^{14} Magnification • 175,000 iterations/second • Frames generated average < 2 min. • TIFF, LaserJet+, etc.





AFFORDABLE PRICING!

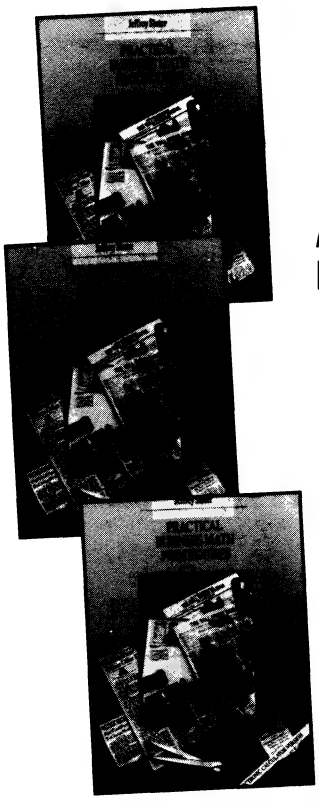
	Software Only	with Source Code
DI-MAN.....	❑ $ 49.95	❑ $ 89.95
DI-MAT.....		❑ $ 49.95
DI-GRAPH	❑ $139.95	❑ $189.95

Shipping/Handling: $5.00 Each Item
TOTAL.....$_____
PA residents add 6% sales tax Make check payable to:

DI-MAT™ $49.95
FAST 32-bit Protected Mode Assembly Matrix Routines

DI-GRAPH™ $139.95
Affordable Hi-res Scientific Plotting Presentation quality • IBM compatible

☞ **DAMASKOS, Inc. Tel. 215-358-0200**
P.O. Box 469 • Concordville,PA 19331 • FAX 215-558-1019



Experience Counts!

New for 1991

Practical Business Math Procedures, 3/e
Slater

Practical Business Math Procedures, Brief Edition
Slater

Practical Business Math Procedures, Brief Edition, Electronic Calculator Version
Slater

Jeff Slater's latest texts tailored to meet your needs.

Also Available in '91

Collegiate Business Mathematics, 5/e
Berston & Fisher

New in '90

Arithmetic Review for Business
Berston & Fisher

The Stancl & Stancl Floor Plan

Mathematics for Management and the Life and Social Sciences
Stancl & Stancl

Brief Calculus for Management and the Life and Social Sciences, 2/e
Stancl & Stancl

Calculus for Management and the Life and Social Sciences, 2/e
Stancl & Stancl

For more information or an examination copy, write or call your IRWIN representative or Faculty Service at 1-800-323-4560. In Canada, call 1-800-268-4178 or 1-416-298-9532.

IRWIN 1818 Ridge Road, Homewood, IL 60430



KEEPING PACE WITH YOUR CHANGING CURRICULUM NEEDS

New and revised titles from PWS-KENT Publishing Company

DEVELOPMENTAL MATHEMATICS

Cass and O'Connor's
Fundamentals with Elements of Algebra: A Bridge to College Mathematics

Proga's
Basic Mathematics, Third Edition

ALGEBRA

Boye, Kavanaugh, and Williams'
Elementary Algebra
Intermediate Algebra

Gobran's
Beginning Algebra, Fifth Edition

GEOMETRY

Geltner and Peterson's
Geometry for College Students, Second Edition

PRECALCULUS

Kaufmann's
Precalculus, Second Edition

CALCULUS

Dick and Patton's
Calculus, Volume 1

Pence's
Calculus Activities for the TI-81

Swokowski's
Calculus, Fifth Edition (formerly the Alternate Edition)
Calculus of a Single Variable

Wiggins'
Problem Solver for Finite Mathematics and Calculus

DISCRETE MATHEMATICS

Fletcher, Hoyle, and Patty's
Foundations of Discrete Mathematics

POST-CALCULUS

Cullen's
Linear Algebra and Differential Equations, Second Edition

ADVANCED MATHEMATICS

Ehrlich's
Fundamental Concepts of Abstract Algebra

Ruckle's
Modern Analysis: Measure Theory and Functional Analysis with Applications

CALL TOLL-FREE ***1-800-343-2204***
(in MA 617-542-3377)



PWS-KENT Publishing Company
20 Park Plaza
Boston, MA 02116
A Division of Wadsworth, Inc.
Partners in Education

Just Published from HBJ!

INTRODUCTORY ALGEBRA
and
INTERMEDIATE ALGEBRA
RONALD HATTON and
GENE R. SELLERS
both of *Sacramento City College*
Each Paperbound, 1991

ELEMENTARY ALGEBRA
Second Edition
and
INTERMEDIATE ALGEBRA
Second Edition
YOSHIKO YAMATO and
MARY JANE CORDÓN
both of *Pasadena City College*
Each Hardcover, 1991

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN,
Drexel University, and
ARNOLD SHAPIRO
Hardcover, 1991

FINITE MATHEMATICS
WILLIAM OWEN and
FRED CUTLIP
both of *Central Washington University*
Hardcover, 1991

When you asked for a sensible, no-frills, mainstream calculus textbook at a reduced cost for your students, we heard you...

CALCULUS
One and Several Variables
ROBERT ELLIS and
DENNY GULICK
both of *University of Maryland*
Hardcover, 1991 $29.95 (Net)!

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of *Vassar College*
Hardcover, 1991

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.,
Michigan State University
Hardcover, 1991

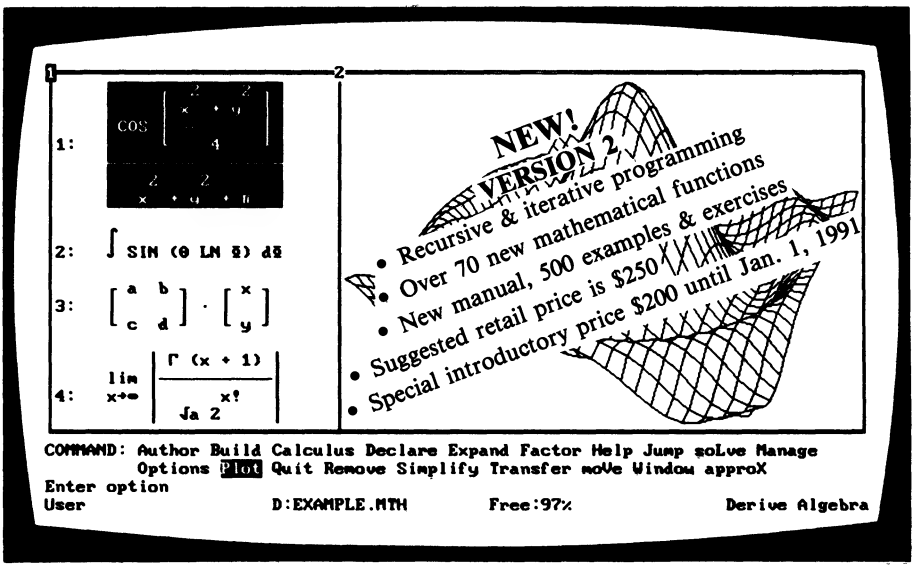
ABSTRACT ALGEBRA
DENNIS KLETZING,
Stetson University
Hardcover, 1991

HBJ

HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, **the successor to muMATH, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:**

- **Exact and approximate arithmetic to thousands of digits**
- **Equations, complex numbers, trigonometry, calculus, vectors, and matrices**
- **2D and 3D function plotting with zooming capability**
- **MDA, CGA, EGA, VGA, and Hercules graphics and text support**
- **Attractive 2D mathematical display of formulas**
- **Easy to use menu-driven interface with on-line help**
- **Ideal for engineers, scientists, students and teachers**
- **$200 plus shipping: Call or write for information.**

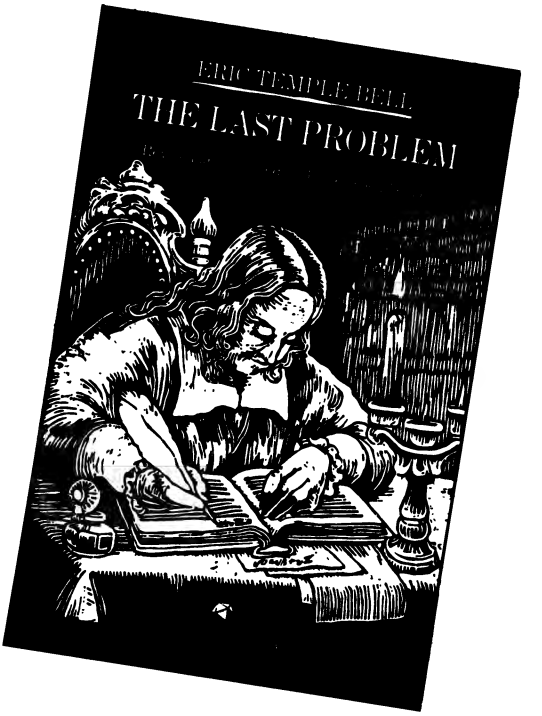
System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. Or NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST
Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.

Revised and Updated



THE LAST PROBLEM

E. T. Bell
Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990
ISBN-0-88385-451-1

List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

Two Books On Mathematical Ingenuity By Ross Honsberger

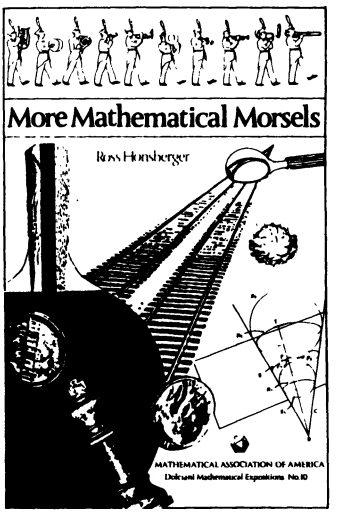
Ross Honsberger is the author of seven books in the Dolciani Mathematical Exposition series, each of which presents problems from algebra, arithmetic, number theory, probability, and geometry, and provides ingenious solutions and/or intriguing results. His most recent addition to the series is MORE MATHEMATICAL MORSELS which is a continuation of the earlier Mathematical Morsels volume published in 1979. Now is your opportunity to own both of these excellent books. The problems presented are meant to be enjoyed, rather than instruct, although instruction is almost always the automatic by-product.

Honsberger says in the Preface of his first volume of morsels, "Mathematics abounds in bright ideas. No matter how long and hard one pursues her, mathematics never seems to run out of exciting surprises. And by no means are these gems to be found only in difficult work at an advanced level. All kinds of simple notions are full of ingenuity. The present volume discusses scores of elementary problems and contains dozens of marvellous ideas."

All of the problems are accessible to anyone with a knowledge of freshman mathematics. Buy both of these excellent books today, and share them with your students.

This is an excellent text...It exposes gifted undergraduate students to a set of elementary problems...whose proofs are ingenious, usually containing beautiful ideas...Its lucid expository style and stimulating mathematical content make its reading a rewarding experience for mathematicians at all levels. (about Mathematical Morsels)

M.S. Cheema in *Mathematical Reviews*



...even the more advanced theoretician will find many of the problems fascinating...

D.T. Swift-Hook in *Physics Education (about Mathematical Morsels)*

NEW NEW NEW NEW NEW NEW

MORE MATHEMATICAL MORSELS

Ross Honsberger

315 pp., Paperbound, 1990,
ISBN 0-88385-313-2

List: $16.00 MAA Member: $13.50

Catalog Number DOL-10

AN MAA BESTSELLER

MATHEMATICAL MORSELS

Ross Honsberger

249 pp., Hardbound, 1979
ISBN 0-88385-303-5

List: $ 28.00 MAA Member: $ 21.00

Catalog Number DOL-03

Both volumes purchased now

List: $40.00 MAA Member: $29.00

Catalog Number DOL-310

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 3 March 1991

Contents

(ISSN 0002-9890)

ARTICLES

How Things Float E. N. GILBERT 201

The Exploding Exponential and Other Chaotic Bursts in Complex Dynamics ROBERT L. DEVANEY AND MARILYN B. DURKIN 217

LETTERS TO THE EDITOR 233

NOTES

Algebraic Characterization of Compact Abelian Groups M. V. DESHPANDE AND SUDHIR R. GHORPADE 235

First-Class Functions G. MYERSON 237

On Series of Partial Fractions P. D. BARRY AND D. J. HURLEY 240

A Note on Dixon's Proof of Cauchy's Integral Theorem PETER A. LOEB 242

THE TEACHING OF MATHEMATICS

The Self-Paced Calculus Program at Berkeley MURRAY H. PROTTER 245

Number Systems With a Complex Base: a Fractal Tool for Teaching Topology DANIEL GOFFINET 249

A Characterization of the Cantor Function DONALD R. CHALICE 255

Parametric Differentiation Revisited L. R. BRAGG 259

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 263

Advanced Problems and Solutions 272

REVIEWS

Basic Hypergeometric Series. By George Gasper and Mizan Rahman GEORGE E. ANDREWS 282

Abstract and Concrete Categories. By Jiří Adámek, Horst Herrlich, and George Strecker GABRIELE CASTELLINI 285

TELEGRAPHIC REVIEWS 288

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $32 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

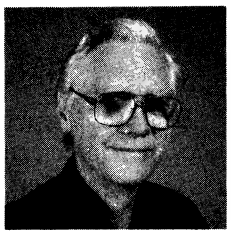
Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

How Things Float

E. N. GILBERT, *AT & T Bell Laboratories, Murray Hill, New Jersey 07974*



EDGAR N. GILBERT received the Ph.D. in mathematics from M.I.T. in 1948. Since then he has been a member of the Mathematics and Statistics Research Center at AT & T Bell Laboratories. His main interests have been communication theory, queueing and other applications of probability, routing problems, and acoustics.

ABSTRACT S. M. Ulam once asked if spheres are the only homogeneous bodies that can float in every orientation. Here an affirmative answer is given for a special class of bodies of revolution that seemed likely to contain counterexamples. Traditionally, the floating bodies of most interest have been ships, which are not homogeneous. However, homogeneous bodies are found to float in surprising ways. Bodies with symmetry often float asymmetrically. Thus, circular cylinders can float with the axis of symmetry tilted away from vertical at strange angles. A cube or regular tetrahedron can float with neither a vertex, an edge center, nor a face center down. Indeed, like the circular cylinder, a cube may have infinitely many indifferently stable floating orientations that are not isolated, but form a one-parameter family.

1. Introduction. The stable orientations of a floating body present a classic problem of hydrostatics. Naval architects use a well-known stability condition (Theorem 1 below) to design ship hulls that float stably under a variety of loading conditions. With the exception of icebergs (Section 4), floating homogeneous bodies have little practical importance. However they have interesting mathematical properties. A simple law connects the stable floating positions of congruent homogeneous bodies of densities ρ and $1 - \rho$ (Theorem 2). Spheres may be the only homogeneous bodies that can float in all orientations (Ulam [**8**]), but that has never been proved or disproved (Section 3). Theorem 4 proves it for a special class of bodies. A homogeneous body with an axis of rotational symmetry always has an equilibrium orientation with the axis vertical, but the equilibrium may be unstable. For example, only the shortest semiaxes of a homogeneous ellipsoid provide stable equilibria (Section 6). Depending on its dimensions and density, a circular cylinder may float with its axis vertical, horizontal, or tilted at some special angle away from vertical (Section 8). Regular polyhedra have many symmetry axes (through vertices, face centers, and edge centers) but, again, these orientations need not be stable. All of them are unstable for a tetrahedron of density $1/2$ and so this tetrahedron floats in a way that is not easily guessed (Section 5). Likewise, for some values of ρ, the cube floats in an asymmetric way. In fact, like many circular cylinders, some cubes do not float in isolated stable orientations; instead they have infinitely many indifferently stable orientations, forming a one-parameter family (Section 10).

2. Potential and metacenter. It is convenient to use coordinates fixed in the body. The problem then is to predict where the waterline will be on the body when it is floating stably. Let n be a unit normal to the water surface, directed up from the water. It suffices to find n since then Archimedes' principle determines the waterline.

Stable positions for a floating body occur at minima of a potential energy function $U(n)$. $U(n)$ is the work done to bring the body into a position with n

upward and with the waterline at the level where the body displaces its own weight in water. The work done against gravity is mgS_1 where m is the mass of the body, g the acceleration of gravity, and S_1 the height of the center of gravity G of the body. The work done against buoyancy forces is just the work required to raise the particles of the displaced water up to the water level. This work is mgS_2 where S_2 is the depth of the centroid B of the submerged part of the body. B is called the *center of buoyancy*. Thus

$$U(n) = mg(S_1 + S_2) = mgn \cdot (G - B). \tag{1}$$

Since the gravity force acts downward from the center of gravity and the buoyancy force acts upward from the center of buoyancy, the line joining B and G must have the vertical direction n in equilibrium. Some equilibrium directions are maxima or saddle points of the potential function $U(n)$ and represent unstable equilibrium orientations of the body.

If n_0 is an equilibrium direction, the behavior of $S(n) \equiv n \cdot (G - B)$ near n_0 determines the type of equilibrium. To perturb n away from n_0, rotate the body through a small angle α about an axis with unit direction vector t normal to n_0. Then $n = n_0 \cos \alpha + (n_0 \times t)\sin \alpha$. If $S(n)$ has continuous partial derivatives of order 3 at n_0,

$$S(n) = S(n_0) + \text{const } \alpha^2 + O(\alpha^3).$$

Buoyancy forces will tend to move the body back to orientation n_0 if the constant in the quadratic term is positive, or further away from n_0 if the constant is negative. The main theorem about floating bodies expresses this constant in terms of a point M, called the metacenter, which is defined as follows.

For a given equilibrium direction n_0, let V denote the volume of the submerged part of the body and let A denote the two-dimensional figure in which the plane of the waterline intersects the body. Let I denote the moment of inertia of A about an axis with direction t ($t \cdot n_0 = 0$) and passing through the centroid of A. Then the *metacenter* for n_0 and t is the point

$$M = B + n_0 I/V \tag{2}$$

lying on the (vertical) line through G and B at height I/V above B.

THEOREM 1. *When n is rotated away from equilibrium direction n_0, through a small angle α about axis t,*

$$n \cdot (G - B) = S(n) = S(n_0) + \frac{n_0 \cdot (M - G)\alpha^2}{2} + O(\alpha^3). \tag{3}$$

Buoyancy forces act to move the body back to orientation n_0 if M lies above G and away from n_0 if M lies below G.

The idea of metacenter and its connection with stability date back to Huyghens. Theorem 1 appears in many books on mechanics ([**1**], [**9**]) and naval architecture ([**6**]).

Since the metacenter depends on t, an equilibrium orientation n_0 may be stable with respect to some perturbation axes t and unstable for others (saddle-point equilibrium). If the principal moments of inertia of A about its centroid are I_1 and I_2, with $I_1 \leqslant I_2$, let the metacenters for rotations about the corresponding principal axes be M_1 and M_2. Then n_0 is a minimum (stable point), saddle point, or maximum of $S(n)$ if G lies below M_1, between M_1 and M_2, or above M_2. In naval architecture M_1 and M_2 are the metacenters for rolling and pitching.

B, M_1, and M_2 are determined entirely by the shape of the submerged part of the body but G may be moved about by redistributing cargo. To make a ship float stably in a particular orientation n_0, its mass must be distributed to put G below M_1. For instance a small boat, originally stable, can become unstable if its passenger stands up. A vessel partly filled with liquid (a tanker, leaky ship, or half-empty bottle) requires a more complicated treatment; its cargo includes liquid that moves as the body is perturbed.

In what follows the body is a solid made from homogeneous material of density $\rho < 1$. This is not an important case for naval architecture but it contains some mathematical curiosities. Symmetry axes of bodies like regular polyhedra, cylinders, or cones are obvious equilibrium directions. The kinds of stability that these bodies exhibit are often surprising.

3. Indifferent stability. Because of its symmetry, a homogeneous sphere has the same potential $U(n) = mgn \cdot (G - B)$ for all orientations n. Every n is then an equilibrium direction. But these equilibria are indifferently stable. That is, the perturbed body tends neither to return to its original orientation nor to move away from it. Likewise a long circular cylinder has the same potential $U(n)$ for all directions n normal to the cylinder's axis. These directions are equilibria, indifferently stable with respect to rolling although stable with respect to pitching.

Ulam [8] asked, "A solid S of uniform density ρ has the property that it will float in equilibrium (without turning) in water in every orientation. Must S be a sphere?" No nonspherical body with this property is known. H. Auerbach [2] settled a similar problem for cylindrical bodies by constructing noncircular cylinders that float indifferently stably with respect to rolling, in the manner of a circular cylinder. FIGURE 1 is the cross section of one of Auerbach's indifferently stable cylinders. Chords drawn in FIGURE 1 represent waterlines for different floating positions. These chords cut the figure in two halves of equal area ($\rho = 1/2$). Different heart-shaped cross sections work for other densities and there are other solutions that are not heart-shaped.

For a body to have constant $U(n)$, as Ulam required, (1) shows $B = G - hn$ with h a constant. As n varies, the center of buoyancy remains on a sphere of

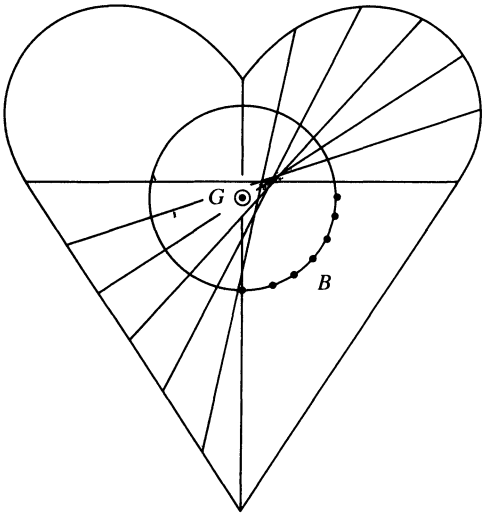


FIG. 1. Cross section of indifferently stable floating cylinder of density 1/2. Chords are possible waterlines. Centers of buoyancy B lie on a circle about G.

radius h and center G. Moreover, (3) requires $M = G$ and then (2) shows $I = hV$, a constant for all n and t.

In Auerbach's cylinder problem, the same reasoning requires that centers of buoyancy lie on a circle, as FIGURE 1 shows. To achieve constant moment of inertia I for rolling, the waterline chords in FIGURE 1 all have the same length. These constraints supply the differential equation that determined the shape of FIGURE 1.

When ρ is small, the body floats with small submerged volume and, in the limit as $\rho \to 0$, as though resting on a flat table. Ulam also asked, "If a body rests in equilibrium in every position on a flat horizontal surface is it a sphere?" The answer to that question must be "yes," at least for bodies that touch the flat surface at only one point $P(n)$. For, as $\rho \to 0$, B must approach $P(n)$; then boundary points $P(n)$ all lie on a sphere. For a purely geometric argument, not using the potential or metacenter, see Montejano [**4**].

4. Capsized bodies. The stable equilibrium positions of a floating body made from homogeneous material of density ρ will ordinarily depend on ρ. If equilibria are known for $\rho \leqslant 1/2$, the following theorem will immediately supply equilibria for $\rho \geqslant 1/2$.

THEOREM 2. *Consider a homogeneous floating body that may have either density ρ or $\rho' = 1 - \rho$. Let the potential functions of the body with densities ρ and ρ' be $U(n)$ and $U'(n)$. Then $U'(-n) = U(n)$.*

Proof. Let the body have volume K. The body with density ρ or ρ' displaces water volumes $V = \rho K$ or $V' = \rho' K$. For a given direction vector n let B denote the center of buoyancy of the body with density ρ. Write B' for the centroid of the part of the body above the water level. This part has volume $K - V = V'$ and so B' is the center of buoyancy when the body with density ρ' is oriented with direction vector $-n$.

From $KG = VB + V'B'$ one finds $V(G - B) + V'(G - B') = 0$ and then

$$-n \cdot (G - B')V' = n \cdot (G - B)V. \tag{4}$$

Now (1) and (4) show that $U'(-n) = U(n)$. ■

An immediate consequence of Theorem 2 is

COROLLARY. *If n_0 is an equilibrium direction for a body of density ρ then $-n_0$ is one for the same body with density $1 - \rho$. The two equilibria are alike (both minima, both maxima, or both saddle points). Bodies with $\rho = 1/2$ float as stably capsized as right-side-up.*

FIGURE 2 illustrates the corollary by showing stable positions for a body with densities ρ and $1 - \rho$.

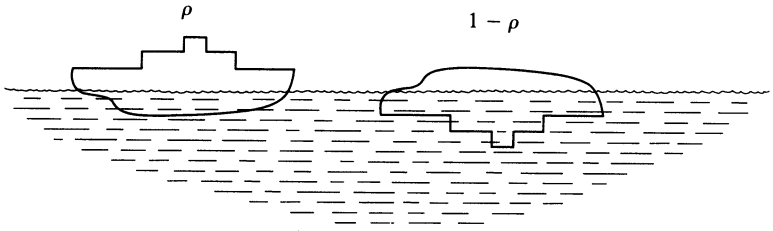


FIG. 2. Stable floating positions for congruent homogeneous bodies of density ρ and $\rho' = 1 - \rho$.

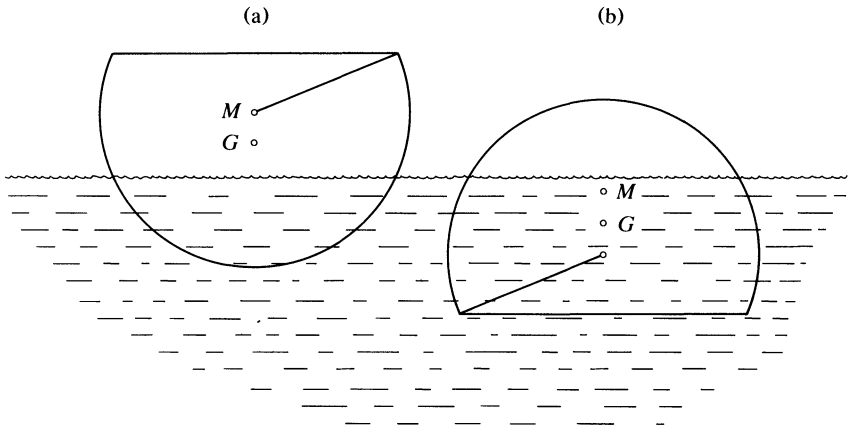


FIG. 3. Two stable orientations of truncated spheres.

For another example, consider a sphere truncated by a plane cut. In one equilibrium position the plane cut is horizontal and above the water level (Figure 3a). The submerged part of the body is spherical and so, as Section 2 remarked, the metacenter M lies at the center of the sphere. Truncating the sphere has moved G from the center M of the sphere to a lower point. Then this equilibrium is stable for any ρ. But now the corollary shows that truncated spheres of all densities also float stably when capsized, with the cut submerged (Figure 3b).

The corollary can help show how icebergs float. Ice has density about $\rho = .9$ relative to sea water. Consider, instead, a fictitious iceberg with density $\rho' = 1 - \rho = .1$. This light body may be expected to float in nearly the orientation it would assume resting on a flat plane. Then the real iceberg floats in the capsized orientation. For example, an iceberg with a large flat face is likely to float face-up. Of course, as Figure 3b illustrates, it may also float in other orientations, although not as stably.

5. Counting equilibria. In looking for equilibrium points a useful identity relating the numbers of maxima, minima, and saddle points is

$$N(\text{max}) + N(\text{min}) = N(\text{saddle}) + 2, \tag{5}$$

where $N(\cdots)$ is read "number of...". Equation (5) is a topological identity that applies to functions $U(n)$ with continuous second derivatives, defined on the sphere (see [5] and [7]). It assumes equilibrium points are isolated and only of the three kinds indicated. Thus, near an equilibrium point, $U(n)$ must agree to terms of second order with a non-degenerate quadratic form in local coordinates.

The identity provides a check that all equilibrium points have been found and properly classified. For example, the directions of the six semiaxes of a homogeneous ellipsoid are obvious equilibrium points. In Section 6, these are shown to be the only equilibria; there are two maxima, two minima, and two saddle points, in agreement with (5).

The identity (5) has an interesting application to the equilibria of a floating regular polyhedron. The directions from the polyhedron's center to the vertices, face centers, and edge centers are obvious equilibrium directions because of symmetry. Indeed, in the limit as $\rho \to 0$, the body can touch the water surface only on a face, edge, or vertex, and these equilibria are the only possibilities. Moreover their numbers satisfy Euler's identity

$$N(\text{faces}) + N(\text{vertices}) = N(\text{edges}) + 2. \tag{6}$$

With the light regular polyhedron resting on the water surface as though on a table, each face, vertex, or edge provides a minimum, maximum, or saddle point of $U(n)$ (this need not be true of irregular polyhedra) and (5) is an instance of (6).

One might expect directions along axes of symmetry to be the only equilibria even when $\rho > 0$. Axes through faces and vertices are axes of 3-fold symmetry or more; then $M_1 = M_2$ for these axes and their equilibria can be only maxima or minima. Axes through edge centers have only 2-fold symmetry and so might give the saddle-point equilibria that (5) demands. However, the situation can be more complicated. A regular tetrahedron with $\rho = 1/2$ provides a quick example. The corollary to Theorem 2 shows that it floats as stably face-down as vertex-down; indeed the metacenters, obtained from (2), show both orientations to be maxima of $U(n)$. Moreover the orientations with edge down also turn out to be maxima. Then other orientations, not along symmetry axes, will minimize $U(n)$ and even more will provide saddle points.

FIGURE 4a will help find the missing equilibria. It is convenient to represent each n by the point where the ray from G, with direction n, crosses a face of the tetrahedron (gnomonic projection). The corollary to Theorem 2 states that equilibrium directions come in pairs, represented by diametrically opposite points, such as vertex P_4 and face center F, or the edge centers E and E'. The plane of P_4, E', P_3, F, E, and G is a plane of mirror symmetry of the tetrahedron. Then maxima and minima of $U(n)$, with n restricted to vectors lying in the plane, occur only at equilibrium points. Since P_3, F, and E are maxima there must also be two minima, Q between P_3 and F, and R between F and E. To locate Q exactly, note that the plane of P_1, E', P_2, E is another plane of mirror symmetry, and hence a waterline

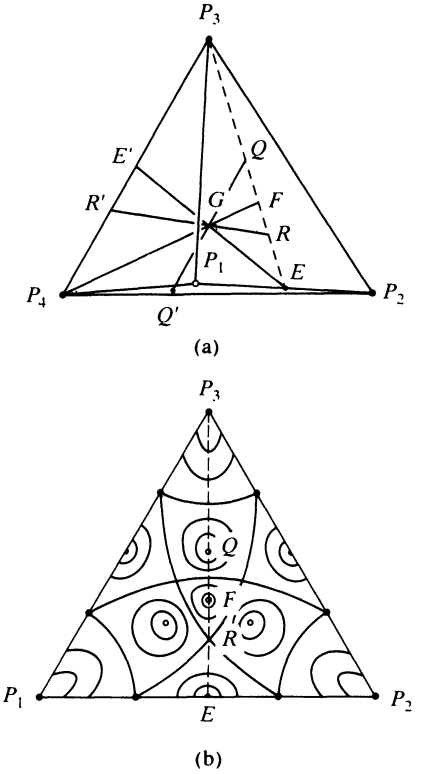


FIG. 4. Equilibrium points and equipotential lines on a regular tetrahedron of density 1/2.

of an equilibrium orientation of the tetrahedron. The normal to this plane is parallel to edge P_4P_3 and represented by Q, the midpoint of the dotted altitude P_3E of face $P_1P_2P_3$. A metacenter calculation shows Q to be a stable point. The point R is more complicated but, lying between F and E, R is diametrically opposite a point R' on edge P_3P_4. Thus the missing equilibria include 12 minima like Q, on midpoints of face-altitudes, 12 points like R, and 12 like R'. Equation (5) will be satisfied if R and R' are saddle points and there are no further equilibria. The laborious calculations, required to locate R and R', find their type of equilibrium, and verify that all equilibria have been found, have not been made. Figure 4b is only an artist's conception of the lines of constant $U(n)$ on a face.

The stable orientations of a tetrahedron of density $1/2$ are related to symmetry because they hold one edge vertical. However, with ρ slightly different from $1/2$, these orientations are no longer equilibria and the body must float asymmetrically.

6. Ellipsoid. The stability of a floating ellipsoid will be determined by first showing how compressing a body affects its stability. Consider any floating body F made of homogeneous material of density ρ. If n_0 is an equilibrium position for F, introduce cartesian body-coordinates x, y, z with the z-axis in the direction n_0. A transformation

$$x' = ax, \qquad y' = by, \qquad z' = cz \tag{7}$$

compresses F into a body F'. Volumes V and V' of corresponding regions in F and F' are related by

$$V' = abcV.$$

If F' is also made of homogeneous material of density ρ, the compression (7) maps the waterline of F to the waterline of F' for the same direction n_0. Centers of gravity G and buoyancy B of F also map to the corresponding centers G' and B' of F'. Then n_0 is also an equilibrium direction of F'. However F and F' may have different types of equilibrium at n_0 because (7) does not map metacenters into metacenters.

Let M and M' be the metacenters of F and F' for small rotations of n about the x-axis. These metacenters are related because the waterline figures A, A' of F, F' have moments of inertia I, I' satisfying $I' = ab^3I$. Then (2) shows

$$n_0 \cdot (M' - B') = I'/V' = ab^3I/(abcV) = \frac{b^2}{c} n_0 \cdot (M - B).$$

The stability condition $0 < n_0 \cdot (M' - G')$ for F' is equivalent to $n_0 \cdot (G' - B') < n_0 \cdot (M' - B')$, or

$$\frac{b^2 n_0 \cdot (M - B)}{c^2 n_0 \cdot (G - B)} > 1. \tag{8}$$

Now, in particular, let F be a unit sphere so that the compression (7) produces an ellipsoid F', with semiaxes a, b, c, floating with c-semiaxis vertical. Because the sphere is indifferently stable, (3) shows that all its metacenters M coincide with G. The condition (8) for stability with respect to small rotations about the a-semiaxis, starting from a configuration with n_0 along a c-semiaxis, reduces to $c < b$. Similarly F' will be stable with respect to small rotations about a b-semiaxis if $c < a$. It follows that the shortest semiaxes of F' are stable directions n, the longest semiaxes maximize the potential $U(n)$, and the middle-sized semiaxes provide saddle points.

If the semiaxes a, b, c of an ellipsoid F' are all different, they provide six equilibrium directions, two each of minima, maxima, and saddle points. Equation (5) suggests that these may be the only equilibrium directions. A proof that there are no other equilibrium directions follows.

Consider the ellipsoid $(x'/a)^2 + (y'/b)^2 + (z'/c)^2 = 1$ obtained from a unit sphere $x^2 + y^2 + z^2 = 1$ by the compression (7). Suppose $a < b < c$. Imagine the ellipsoid floating with upward direction $n' = (\lambda', \mu', \nu')$; let $\lambda'x' + \mu'y' + \nu'z' = d'$ denote the water surface. On the unit sphere the corresponding water surface is $\lambda x + \mu y + \nu z = d$ where $n = (\lambda, \mu, \nu)$ is a unit vector with the direction of $(\lambda'a, \mu'b, \nu'c)$. The center of buoyancy B of the sphere lies below the origin G, in the direction of $-n$, say $B = -(\beta\lambda'a, \beta\mu'b, \mu\nu'c)$. Since affine transformations carry centroids into centroids, the center of buoyancy of the ellipsoid is at

$$B' = -(\beta a^2\lambda', \beta b^2\mu', \beta c^2\nu').$$

In order for n' to be an equilibrium direction of the ellipsoid, n' and B' must have the same direction. Since a^2, b^2, c^2 are all different, the only possibilities for n' are $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, $(0, 0, \pm 1)$, the semiaxis directions.

7. Body of revolution. For bodies of revolution, $U(n)$ simplifies to a function $u(\theta)$ of just the angle θ between n and the axis a of revolution $(0 \leqslant \theta \leqslant \pi)$. The axial directions $n = \pm a$ are obvious equilibria. Other equilibrium directions n are not isolated, but lie on circles of constant θ. Then (5) does not apply.

For equilibria other than $n = \pm a$ it is convenient to use the term *pitch* for a rotation about a horizontal axis p that is normal to a. Then a *roll* is a rotation about a horizontal axis normal to p, i.e., an axis in the plane of n and a. Because the plane of n and a is a plane of mirror symmetry of the waterline figure A, the pitch and roll axes are principal directions for A. Then M_1 and M_2 are the metacenters for pitching and rolling (although perhaps not in that order).

THEOREM 3. *Consider a homogeneous floating body of revolution with axial direction vector a. At an equilibrium direction $n_0 \neq \pm a$, the roll metacenter M_r is at the center of gravity G.*

Proof. Introduce body coordinates x, y, z with z axis along the axis of rotation and x-axis in the plane of a and n_0. Write $n_0 \cdot a = \cos\theta_0$. Then $n_0 = (\sin\theta_0, 0, \cos\theta_0)$. A roll is a rotation about the x-axis. A roll of size α moves the vertical direction from n_0 to $n = (\sin\theta_0, \cos\theta_0 \sin\alpha, \cos\theta_0 \cos\alpha)$. Then $n \cdot a = \cos\theta = \cos\theta_0 \cos\alpha$ and, for small α, $\theta = \theta_0 + (\alpha^2/2)\cot\theta_0 + O(\alpha^4)$. The potential at n is

$$U(n) = u(\theta) = u(\theta_0) + \frac{du}{d\theta}(\theta - \theta_0) + O(|\theta - \theta_0|^2).$$

At n_0, $du/d\theta = 0$ because n_0 is an equilibrium direction. Then $U(n) = U(n_0) + O(|\theta - \theta_0|^2) = U(n_0) + O(\alpha^4)$ and (3) shows $M_r = G$. ■

Theorem 3 shows that a body of revolution is indifferently stable with respect to small roll perturbations. This result can also be deduced from the observation that the body is indifferently stable to rotations about its axis a of symmetry. For the sake of simplicity, an equilibrium of a body of revolution will be called stable if it is pitch-stable.

COROLLARY. *An equilibrium $n_0 \neq \pm a$ of a body of revolution is stable iff I_p and I_r, the moments of inertia of A for pitching and rolling, satisfy $I_r < I_p$.*

The corollary follows immediately from (2) and Theorems 1 and 3.

An ellipsoid with two semiaxes equal is an ellipsoid of revolution, a spheroid. The results of Section 6 show that the rotation axis a must be either vertical ($\theta = 0$ or π) or horizontal ($\theta = \pi/2$) in any equilibrium position of a spheroid. Oblate spheroids are stable with a vertical and prolate spheroids are stable with a horizontal. Other bodies of revolution can have more complicated equilibria. The circular cylinder of the next section is a typical example.

8. Cylinder. In this section the body is a circular cylinder of radius R and height H.

In one equilibrium position the cylinder floats on end, with $\theta = 0$ or $\theta = \pi$ in the notation of Section 7. A simple calculation of the pitch metacenter shows that this orientation is stable if

$$\rho(1-\rho)\left(\frac{H}{R}\right)^2 < \frac{1}{2}. \tag{9}$$

In another equilibrium the cylinder floats on its side, with $\theta = \pi/2$. The waterline figure A is a rectangle, with sides H and w, where Archimedes' principle determines w as a root of

$$\arcsin\frac{w}{2R} - \frac{w}{2R}\sqrt{1-\left(\frac{w}{2R}\right)^2} = \rho\pi.$$

The corollary to Theorem 3 shows that this equilibrium is stable if

$$w < H. \tag{10}$$

Both (9) and (10) may hold if ρ is small, only (9) if H/R is small, only (10) if H/R is large, and for some parameter values (e.g., $\rho = 1/2$, $H/R = 1.7$) neither (9) nor (10) hold. Then some cylinders must float with axis tilted at other angles θ, not 0, $\pi/2$, or π. There are several cases to consider because the waterline may intersect either end of the cylinder, neither, or both. In what follows x, y, z are cartesian coordinates, fixed in the body, with the z-axis placed along the axis of revolution. Then the cylinder is $x^2 + y^2 \leqslant R^2$, $0 \leqslant z \leqslant H$.

In the simplest configuration, the end $z = 0$ is completely submerged and the end $z = H$ is completely out of the water. The waterline plane passes through the points $(-R, 0, \rho H - R\tan\theta)$, $(0, 0, \rho H)$, and $(R, 0, \rho H + R\tan\theta)$. This configuration is possible only if

$$\frac{R}{H}\tan\theta \leqslant \operatorname{Min}\{\rho, 1-\rho\}. \tag{11}$$

If this orientation is an equilibrium, it is stable. For, the waterline figure A is an ellipse with semiaxes R and $R\sec\theta > R$; the corollary to Theorem 3 ensures pitch-stability. The equilibrium condition is that G be directly above B, i.e., the line GB must make angle θ with the axis of the cylinder. Now $G = (0, 0, H/2)$ and $B = (B_x, 0, B_z)$ where

$$B_x = \frac{R^2\tan\theta}{4\rho H}, \quad B_z = \frac{\rho H}{2}\left\{1 + \left(\frac{R\tan\theta}{2\rho H}\right)^2\right\}.$$

The condition for stability is $B_x = (H/2 - B_z)\tan\theta$, or

$$\rho(1-\rho)\left(\frac{H}{R}\right)^2 = \frac{1}{2} + \frac{1}{4}\tan^2\theta. \tag{12}$$

If $\rho(1-\rho)(H/R)^2 < 1/2$, there is no equilibrium of this kind, but then (9) shows that the cylinder floats stably on its end. For larger values, (12) determines an angle θ. If θ satisfies (11), then the cylinder floats stably tilted at angle θ. If (11) is violated, (12) no longer applies and the equilibrium is one with waterline passing through one or both ends.

To simplify the problem, only the case $\rho = 1/2$ will be considered. Formula (12) with $\rho = 1/2$ determines an angle θ satisfying (11) if $2 < (H/R)^2 < 8/3$. When $H/R = \sqrt{8/3} = 1.63299$, the tilt angle is $\theta = 39.23°$ and the two ends of the cylinder are just tangent to the water surface. For larger values of H/R, consider a configuration with waterline plane passing through $(-(H/2)\cot\theta, 0, 0)$, $(0, 0, H/2)$, and $((H/2)\cot\theta, 0, H)$. The formulas that follow are simplified by the notation

$$t = \frac{H}{2R}\cot\theta, \quad J_k = \int_0^t x^k\sqrt{1-x^2}\,dx \tag{13}$$

for $k = 0, 2$. The integrals J_k are elementary,

$$J_0 = \frac{1}{2}\{t\sqrt{1-t^2} + \arcsin t\}$$

$$J_2 = \frac{1}{4}\{J_0 - t(1-t^2)^{3/2}\}.$$

The waterline in this configuration passes through both faces. That requires $(H/2)\cot\theta < R$, i.e., $t < 1$.

Now $V = \pi R^2 H/2$ and $B = (B_x, 0, B_z)$ with

$$VB_x = \frac{4}{3}R^4(J_0 - J_2)$$

$$VB_z = R^2H^2\left(\frac{\pi}{4} - \frac{1}{2}J_0\right) + 2R^4\tan^2\theta J_2.$$

The equilibrium condition

$$B_x = \left(\frac{H}{2} - B_z\right)\tan\theta$$

is

$$\tan^2\theta = \frac{2}{3}\frac{J_0 - J_2}{J_0t^2 - J_2}. \tag{14}$$

The stability of this equilibrium must be determined by finding the principle moments of inertia of the waterline figure A. A is no longer the entire ellipse, with semiaxes R and $R\sec\theta$, because that ellipse protrudes beyond the ends of the cylinder. To get A, pieces at the ends of the ellipse must be cut off to leave a figure of length $H\csc\theta$ (see FIGURE 5). The moments for pitching and rolling are

$$I_p = 4R^4\sec^3\theta J_2$$

$$I_r = \tfrac{4}{3}R^4\sec\theta(J_0 - J_2).$$

The corollary to Theorem 3 now gives the stability condition $I_p > I_r$, or

$$\cos^2\theta < \frac{3J_2}{J_0 - J_2}. \tag{15}$$

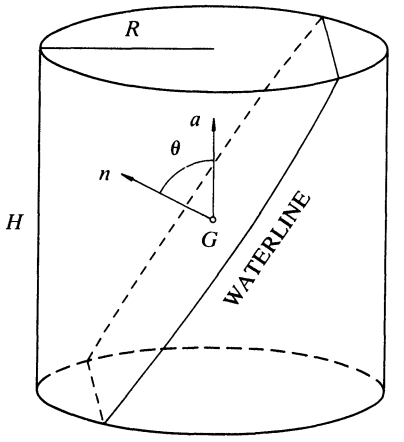


FIG. 5. Waterline figure A of floating circular cylinder.

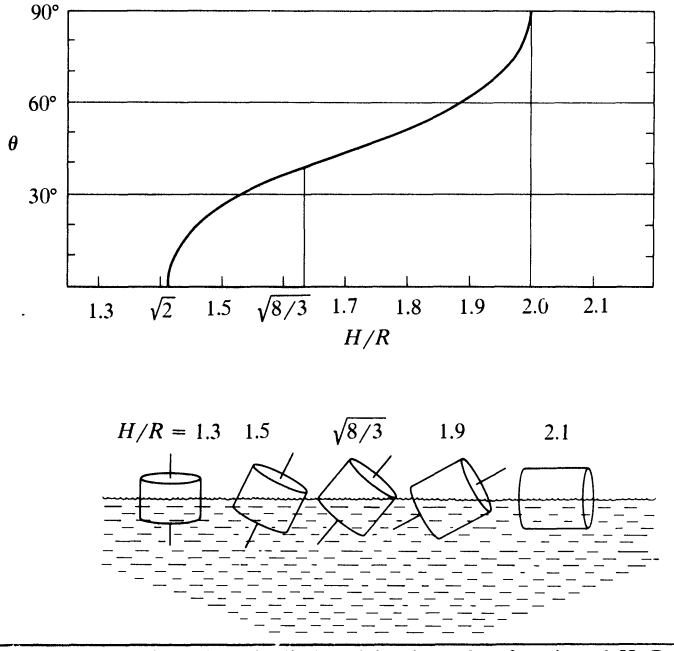


FIG. 6. Stable orientation of cylinder of density 1/2 as function of H/R.

Each value of t in $0 < t < 1$ determines an angle θ in (14) and then (13) gives $H/R = 2t \tan \theta$. A check on the condition (15) shows that the equilibrium is stable throughout $0 < t < 1$. As t varies, H/R and θ range over $2 > H/R > \sqrt{8/3}$ and $90° > \theta > 39.23°$.

The cylinder with density $\rho = 1/2$ has, for each H/R, a unique angle θ at which it floats stably. FIGURE 6 shows how the stable orientation varies with H/R. Formulas $\theta = 0°$, (12), (14), and $\theta = 90°$ hold in the ranges $0 < H/R < \sqrt{2}$, $\sqrt{2} < H/R < \sqrt{8/3}$, $\sqrt{8/3} < H/R < 2$ and $2 < H/R < \infty$.

9. Inversion symmetry. A body has *inversion symmetry* if, when the origin O of coordinates is placed at the centroid G, the inversion $(x, y, z) \to (-x, -y, -z)$ leaves the body unchanged. If a body has inversion symmetry, then any plane

through G cuts the body into two congruent halves. If a body has $\rho = 1/2$, as well as inversion symmetry, then its waterlines are easy to find. The waterline, for any given n, lies in the plane normal to n that passes through G.

Bodies of revolution, being already indifferently stable with respect to rolling, seem likely candidates to be indifferently stable to all perturbations, as requested by Ulam.

THEOREM 4. *Except for spheres, there are no indifferently stable bodies of revolution with $\rho = 1/2$ that also have inversion symmetry.*

Proof. Introduce spherical polar coordinates (r, γ, δ) with origin at G and with polar angle γ (the axis of revolution is $\gamma = 0$ and $\gamma = \pi$). An equation of form $r = r(\gamma)$ will describe the surface of the body.

Consider the body floating with upward direction vector n and let n have angle coordinates $\gamma = \theta$, $\delta = 0$. Now introduce a second system of spherical polar coordinates (r, α, β) with pole $\alpha = 0$ at n. The original polar angle γ is related to the new coordinates by

$$\cos\gamma = \cos\theta\cos\alpha + \sin\theta\sin\alpha\cos\beta. \tag{16}$$

The potential $U(n)$ is (1) with

$$S(n) = -\iiint (P \cdot n) r^2 \sin\alpha \, dr \, d\alpha \, d\beta,$$

wherein the integration extends over points P of the body below the waterline plane. The range of integration is

$$0 \leqslant r \leqslant r(\gamma), \quad 0 \leqslant \beta \leqslant 2\pi, \quad \frac{\pi}{2} \leqslant \alpha \leqslant \pi,$$

where γ depends on α and β, as in (16). Because $P \cdot n = r\cos\alpha$, the r-integration is immediate;

$$S(n) = -\frac{1}{4}\iint r^4(\gamma)\cos\alpha\sin\alpha \, d\alpha \, d\beta. \tag{17}$$

Expand $r^4(\gamma)$ in a Neumann series

$$r^4(\gamma) = \sum A_k P_k(\cos\gamma) \tag{18}$$

with $P_k(\cdot)$ the kth Legendre polynomial. Because the body has inversion symmetry, $r^4(\pi - \gamma) = r^4(\gamma)$, and coefficients A_k with odd k vanish. With $\cos\gamma$ given by (16) the addition theorem for Legendre polynomials ([**3**]) expresses $P_k(\cos\gamma)$ in terms of Legendre polynomials $P_k(\cdot)$ and associated Legendre polynomials $P_k^m(\cdot)$ of $\cos\alpha$ and $\cos\theta$

$$P_k(\cos\gamma) = P_k(\cos\alpha)P_k(\cos\theta) + 2\sum_{m=1}^{k} \frac{(k-m)!}{(k+m)!} P_k^m(\cos\theta)P_k^m(\cos\alpha)\cos m\beta.$$

Then when (18) is substituted into (17), the integration on β removes the terms containing associated Legendre polynomials, and leaves

$$S(n) = -\frac{\pi}{2}\left\{A_0 + \sum A_k P_k(\cos\theta)\int_0^1 P_k(x)\,x\,dx\right\}. \tag{19}$$

Now suppose the body is indifferently stable, so that $S(n)$ is constant. In (19), the coefficient of each term $P_k(\cos\theta)$ must vanish. With k even, say $k = 2r \geqslant 2$,

$$\int_0^1 P_k(x)\,x\,dx = \frac{(-1)^{r+1}}{4^r(2r-1)(2r+1)}\binom{2r}{r} \neq 0$$

and so $A_k = 0$. Also, as was mentioned above, $A_k = 0$ for odd k because of the inversion symmetry. Then (18) becomes $r^4(\gamma) = A_0$ and the body is a sphere. ■

Bodies with inversion symmetry are so special that the theorem does not strongly suggest that the sphere may be the only indifferently stable body. For example, in the problem of indifferently roll-stable cylinders, a very similar argument would show there are no non-circular examples with inversion symmetry. Yet, Auerbach produced examples without inversion symmetry.

10. Cube. This section will show that a homogeneous cube can float in surprising ways. Again, x, y, z will be body coordinates, now in the interval $[0, 1]$. Then $G = (1/2, 1/2, 1/2)$. Three planes through G, each normal to a coordinate axis, are planes of mirror symmetry of the cube. It follows, with $n = (\lambda, \mu, \nu)$, that $U(n)$ depends only on the magnitudes $|\lambda|, |\mu|, |\nu|$. In what follows λ, μ, ν will be assumed non-negative. It will also suffice to restrict ρ to $0 \leqslant \rho \leqslant 1/2$ because Theorem 2 will then show how denser cubes behave. Equilibria with n along axes of symmetry will be considered first; the metacenters for these equilibria are easily calculated.

When the cube floats face-down, $n \cdot (G - B) = (1 - \rho)/2$. Also $n \cdot (M - B) = 1/(12\rho)$ for rotations about any axis t. Then the cube floats stably on its face if $\rho(1 - \rho) < 1/6$, i.e., if $\rho < .211325$.

When the cube floats edge-down,

$$n \cdot (G - B) = \frac{1}{\sqrt{2}} - \frac{2}{3}\sqrt{\rho}\,.$$

The metacenter now depends on the axis t. It will be convenient to use the terms roll and pitch as though the cube were a ship with the submerged edge as its keel. The metacenters M_r, M_p for rolling and pitching satisfy

$$n \cdot (M_r - B) = \frac{2}{3}\sqrt{\rho}\,, \quad n \cdot (M_p - B) = \frac{1}{6\sqrt{\rho}}\,.$$

For $0 \leqslant \rho < 1/8$ the equilibrium is a saddle point of $U(n)$, roll-unstable and pitch-stable. For $1/8 < \rho < 9/32$ it is a maximum. For $9/32 < \rho \leqslant 1/2$ it is again a saddle point, but roll-stable and pitch-unstable.

When the cube floats vertex-down the submerged part of the body is a triangular pyramid for $0 < \rho \leqslant 1/6$. But, when $1/6 < \rho < 1/2$ the waterline plane cuts all 6 faces of the cube and the figure A is an irregular hexagon. The calculations of B and M are longer, although still elementary. Floating vertex-down turns out to be unstable (a maximum) for $0 \leqslant \rho < 1/6$ and stable for $1/6 < \rho \leqslant 1/2$.

Equation (5) shows that axes of symmetry need not be the only equilibrium directions. When $1/8 < \rho < 9/32$ these axes provide no saddle points. A curious phenomenon occurs when $.211325 < \rho < .224602$. In that range the potential $U(n)$ of the unstable face-down orientation is actually lower than the potential of the stable vertex-down orientation; then there must be other stable orientations with even lower potential.

To find the missing equilibria numerically one can just plot $U(n)$ as a function of n on the unit sphere. The function $U(n)$, which will be derived next, can only be piecewise analytic; its form depends on which vertices are submerged.

For fixed $n = (\lambda, \mu, \nu)$, the derivation introduces an integral

$$F(s) = \iiint e^{-(\lambda x + \mu y + \nu z)s}\, dx\, dy\, dz$$

taken over the entire body. For the unit cube,

$$F(s) = \frac{1 - e^{-\lambda s}}{\lambda s} \frac{1 - e^{-\mu s}}{\mu s} \frac{1 - e^{-\nu s}}{\nu s}, \tag{20}$$

but the same argument will apply to a general body as long as the origin is located so that the body lies entirely in the positive octant.

THEOREM 5. *For $k = 1, 2$, write $f_k(h)$ for the inverse Laplace transform of $F(s)/s^k$. When the water surface is the plane $\lambda x + \mu y + \nu z = h$, $f_1(h)$ is the displaced volume V and $f_2(h)/V$ is the depth of the center of buoyancy B below the water level.*

Proof. Write $A(h)$ for the area of intersection of the plane $\lambda x + \mu y + \nu z = h$ with the body. Then

$$F(s) = \iiint e^{-hs}\, dx\, dy\, dz = \int_0^\infty e^{-hs} A(h)\, dh,$$

the Laplace transform of $A(h)$. $F(s)/s$ is the inverse Laplace transform of

$$f_1(h) = \int_0^h A(h')\, dh',$$

but that integral is just the displaced volume V below the plane $\lambda x + \mu y + \nu z = h$.

Likewise, $F(s)/s^2$ is the Laplace transform of

$$\begin{aligned} f_2(h) &= \int_0^h \int_0^{h'} A(h'')\, dh''\, dh' \\ &= \int_0^h (h - h'') A(h'')\, dh'', \end{aligned}$$

the second line following by interchanging the order of integration. But this integral is just V times the depth of the center of buoyancy. ∎

In general, Archimedes' principle specifies V. Then the equation

$$f_1(h) = V, \tag{21}$$

from Theorem 5, determines h and the depth $f_2(h)/V$ of B. G is at height $n \cdot G$ above the origin, or $n \cdot G - h$ above the water level. Then the potential $U(n)$ in (1) is proportional to

$$n \cdot (G - B) = n \cdot G + \frac{f_2(h)}{V} - h. \tag{22}$$

For the unit cube, $V = \rho$ in (21) and $n \cdot G = (\lambda + \mu + \nu)/2$ in (22). To find $f_k(h)$, multiply out (20) into a sum of 8 terms, each an exponential divided by $\lambda\mu\nu s^{k+3}$. Transforming this expression term-by-term produces

$$\begin{aligned} f_k(h) = \frac{1}{\lambda\mu\nu(k+2)!} \{ &\|h\|^{k+2} - \|h - \lambda\|^{k+2} - \|h - \mu\|^{k+2} - \|h - \nu\|^{k+2} \\ &+ \|h - \lambda - \mu\|^{k+2} + \|h - \mu - \nu\|^{k+2} + \|h - \nu - \lambda\|^{k+2} \\ &- \|h - \lambda - \mu - \nu\|^{k+2} \}, \end{aligned} \tag{23}$$

where $\|z\|$ means z if $z \geqslant 0$ and 0 if $z < 0$. Each term in (23) represents a different vertex of the cube. Thus h is the depth of $(0, 0, 0)$, $h - \lambda$ is the depth of $(1, 0, 0)$, $h - \lambda - \mu$ is the depth of $(1, 1, 0)$, etc. Only the submerged vertices contribute non-zero terms to (23).

Formulas (21), (22), (23) were used in a numerical study of a cube of density $\rho = .217$. This density lies in the interesting range, discussed earlier, in which the global minimum of $U(n)$ occurs away from the cube's symmetry axes and so remains to be found. FIGURE 7 shows level contours of $n \cdot (G - B)$, drawn in gnomonic projection from G onto a typical face of the cube as in FIGURE 4b. Beside the equilibria at the face center, edge centers, and vertices, there are four more saddle points, one near each corner of the face. Note that equipotentials in Figure 7 do not meet the face edges at right angles because the gnomonic projection is not conformal.

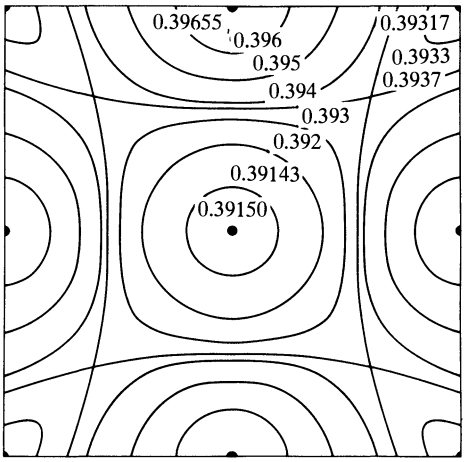


FIG. 7. Equipotential lines on a face of a cube of density $\rho = .217$.

The missing global minima turns out not to be isolated points. They form a circle, with $n \cdot (G - B) = .39143$, surrounding the face center. To show this analytically, consider directions n near $(0, 0, 1)$, a face-center axis. Suppose λ and μ small enough (and $\nu = \sqrt{1 - \lambda^2 - \mu^2}$ large enough) so that the submerged vertices are $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, and $(1, 1, 0)$. Then $f_k(h)$ in (23) is a sum of four terms. When the terms of (23) are expanded in powers of h, considerable cancellation occurs, leaving

$$f_1(h) = \left(h - \frac{\lambda + \mu}{2}\right)\Big/\nu$$

and

$$f_2(h) = \frac{1}{12\nu}\{6h^2 - 6h(\lambda + \mu) + 2\lambda^2 + 3\lambda\mu + 2\mu^2\}. \tag{24}$$

The condition $f_1(h) = V = \rho$ of Theorem 5 provides

$$h = \frac{\lambda + \mu}{2} + \rho\nu. \tag{25}$$

Now substitute (24) into (22) and use (25) to eliminate h. The result, with the help of $\lambda^2 + \mu^2 + \nu^2 = 1$, simplifies to

$$n \cdot (G - B) = \frac{1}{24\rho}\left\{\frac{1}{\nu} + \nu[12\rho(1 - \rho) - 1]\right\} \tag{26}$$

Only the ν component of n enters into (26) and so the level contours of $n \cdot (G - B)$ near the face center are circles. If $\rho(1 - \rho) > 1/6$ (recall that this was also the condition that the face center be an unstable equilibrium point), then $n \cdot (G - B)$ in (26) is a minimum at

$$\nu = \frac{1}{\sqrt{12\rho(1 - \rho) - 1}}.$$

With $\rho = .217$ the minimum occurs at $\nu = .981085 = \cos 11.16°$. For all such n, only the vertices of one face are submerged and (24) does apply.

A cube of density $\rho = .217$ will float with a face-center tilted 11.16° away from vertical. However, the coordinates λ and μ of n are restricted only by $\lambda^2 + \mu^2 = 1 - \nu^2 = .037473$ and the slightest disturbance will move the cube to a new equally stable orientation with the same ν. This cube resembles the cylinder in having a one-parameter family of indifferently stable orientations. In fact, as n ranges over all directions, $n \cdot (G - B)$ remains within 0.65% of .3940 and so the cube comes close to satisfying the condition $n \cdot (G - B) = \text{const}$ of an indifferently stable body.

REFERENCES

1. Paul Appell, Traité de Mécanique Rationelle t3, Gauthier-Villars, Paris (1921), Chapter XXXI, Sec. IV, pp. 194–233.
2. H. Auerbach, Sur un probléme de M. Ulam concernant l'équilibre des corps flottants, *Studia Math*, 7 (1938) 121–142.
3. E. T. Copson, An Introduction to the Theory of Functions of a Complex Variable, Oxford U. Press, London, 1935.
4. L. Montejano, On a problem of Ulam concerning a characterization of the sphere, *Studies in Appl. Math.*, 53 (1974) 243–248.
5. Marston Morse, The Calculus of Variations in the Large, Am. Math. Soc. Colloquium Publ. XVIII, AMS, Providence (1943), Theorem 1.1, p. 143.
6. Henry E. Rossell and Lawrence B. Chapman, Editors, Principles of Naval Architecture, Vol. I, Soc. Naval Architecture and Marine Engineering, New York, 1941, Chapters III and IV.
7. H. Seifert and W. Threlfall, Variationsrechnung im Grossen, reprint Chelsea, New York, 1948, p. 6.
8. S. M. Ulam, A Collection of Mathematical Problems, Interscience, New York, 1960, p. 38.
9. A. G. Webster, The Dynamics of Particles and of Rigid, Elastic, and Fluid Bodies, 2nd ed., G. E. Stechert, New York, 1922.

The Exploding Exponential and Other Chaotic Bursts in Complex Dynamics

ROBERT L. DEVANEY and MARILYN B. DURKIN,* *Boston University*



ROBERT L. DEVANEY did his undergraduate work at Holy Cross and received his Ph.D. from the University of California at Berkeley in 1973. He taught at Northwestern University, Tufts University, and the University of Maryland before coming to Boston University, where he served as chair of the mathematics department from 1983 to 1986. His research is centered in Dynamical Systems, with special interests in Hamiltonian mechanics, complex analytic dynamics, and computer experimentation.



MARILYN B. DURKIN received her B.S. in mathematics from Bates College, her M.S. from Salem State College, and completed her Ph.D. at Boston University in 1989. Since completing her undergraduate work she has held a variety of teaching and private industry consulting jobs. Prior to coming to Bentley College in 1989 she taught at Boston University in the Science and Engineering Program. Her research interests center on Complex Dynamical Systems with a strong emphasis on computer experimentation and the integration of dynamical systems into the undergraduate curriculum.

Introduction. The iteration of complex analytic functions is one of the most alluring subfields of non-linear dynamical systems. One reason for this is the striking beauty of the computer graphics images of Mandelbrot [**Ma**], Peitgen/Richter [**PR**], Douady and Hubbard [**DH**, **DH1**], and others. But, for mathematicians at least, the attractiveness goes much deeper: the combination of powerful techniques from complex analysis and dynamical systems yields a field which has an intrinsic mathematical beauty of its own.

Basically, the problem in complex analytic dynamics, as in many dynamical systems, is to understand the fate of all points under iteration of a map F. That is, given $z \in \mathbf{C}$, what can we say about the sequence of points $z, F(z), F^2(z) = F(F(z)), \ldots, F^n(z), \ldots$ called the *orbit* of z. Here we emphasize that F^n means the n-fold composition of F with itself, not the nth power of F.

There are basically two types of orbits in any dynamical system: those that behave stably or tamely, and those that behave unstably or chaotically. To be imprecise for the moment, an orbit is chaotic if there are nearby points whose orbits tend to diverge far from the given orbit. This means that a chaotic orbit does not lend itself to numerical computation: any small error due to round-off or other numerical approximation may throw us off the original orbit and onto another one whose fate is radically different. The stable orbits, on the other hand, behave more tamely: an orbit is stable if nearby points have orbits which remain nearby for all iterations.

In complex dynamics, the set of all chaotic orbits is called the Julia set, after the French mathematician Gaston Julia who, along with Fatou, first studied the iteration of complex analytic functions in the 1920's [**J**], [**F**]. The Julia set consists of all points that have the property that, if U is any neighborhood of z, then any

*Currently affiliated with Bentley College, Waltham, MA.

point z^* in $\mathbf{C}$ (with at most one exception) has at least one preimage in U for some F^n. That is, the iterates of F tend to smear arbitrarily small neighborhoods of points in the Julia set over almost the entire complex plane. In this sense $J(F)$ is a "strange" repeller. It is a fact that the Julia set contains all of the "interesting" orbits of a complex dynamical system. For this reason, its structure is the subject of much contemporary research.

In this paper we will concentrate on the Julia sets of entire transcendental functions like $\lambda \exp z$ where $\lambda > 0$ is a real parameter. Entire functions share many of the dynamical properties of polynomials or rational maps **[B]**, **[DH]**, but there are some major differences. For example, the Julia sets of these maps have the interesting property that they may explode as the parameter λ is varied. That is, the Julia set undergoes sudden, dramatic changes at certain critical parameter values. In the exponential case, one such explosion occurs when $\lambda = 1/e$. For λ-values less than this critical value, we will show that the Julia set occupies a relatively small, nowhere dense subset of the right half plane, but when $\lambda > 1/e$, the Julia set is the whole plane! This means that the set of chaotic orbits changes discontinuously at this parameter value. This type of bifurcation cannot happen in families of polynomials or rational maps. One of our major goals in this paper is to describe this explosion in detail.

In §1 of this paper we describe some computer experiments which foreshadowed these results. In §2 we present some preliminary notions from complex analysis and dynamical systems theory. We also give a precise definition of the Julia set and formulate some of its basic properties in this section. The goal of §3 is to describe the Julia set of λe^z before the explosion occurs. We prove here that the Julia set is a nowhere dense subset of the right half plane when $\lambda < 1/e$. In §4 we concentrate on the case $\lambda > 1/e$ where the Julia set is the whole plane. We also present several alternative characterizations of the Julia set in these two sections, first as the closure of the set of repelling periodic cycles and second as the closure of the set of escaping orbits. This last characterization gives an efficient algorithm for plotting the Julia sets of these functions. Finally, in §5, we indicate how similar analysis may be used to describe explosions in other families of entire transcendental functions.

1. Computer Experiments. The explosions in the Julia set of an entire function are easy to discover experimentally using a computer. The reason for this is that there is an alternative characterization of the set of chaotic orbits of entire functions which makes computation of the Julia set quite easy. Later in this paper we will show that the Julia set of λe^z may also be defined as the closure of the set of points whose orbits tend to ∞ in the right half plane. While at this stage there appears to be nothing chaotic about an orbit tending to ∞, we will see later that the essential singularity at ∞ for λe^z introduces a tremendous amount of chaotic behavior into the dynamics nearby.

Using this alternative characterization of the Julia set, the algorithm to compute pictures is quite simple:

1. Select a square in the plane and construct a $k \times k$ grid in this square.
2. For each grid point, compute the orbit of this point up to a maximum of N iterations.
3. If, at iteration $i < N$, the real part of the orbit is larger than some bound, B, stop iterating and color the original grid point black.
4. If no point in the orbit ever enters $\operatorname{Re} z > B$, leave the original point white.

In the output generated by this algorithm, we think of the black points as constituting the Julia set and the white points as lying in the complement of the Julia set. This is not quite accurate, as we discuss below, but serves as a good guideline for experimental observations.

In FIGURE 1, we have displayed the Julia sets for several exponentials computed using this algorithm. In each case we have used the bound $B = 50$. One might object that escaping past $\operatorname{Re} z = 50$ hardly constitutes escaping to ∞. However, if $\operatorname{Re} z > 50$, then $|\lambda e^z| > |\lambda| e^{50}$ which is quite large. Moreover, as we will see below, the pictures we obtain with this bound are quite accurate.

FIGURE 1a displays the Julia set of λe^z with $\lambda = .3 < 1/e$. Note that there is a large white area in this figure. The points in this area do not lie in the Julia set. This figure was generated using $N = 25$, but the reader may check that the same figure is obtained using any larger N.

FIGURE 1b displays the Julia set of λe^z with $\lambda = .4 > 1/e$. Note the dramatic change in the size of the Julia set for this value of λ! This is the explosion in the exponential family.

FIGURE 1b was computed using $N = 25$. If we increase N to 50, we obtain FIGURE 1c. Note that the black region has apparently increased in size. Actually, as we discuss below, there should be no white whatsoever in these pictures: the Julia set of λe^z when $\lambda > 1/e$ is the entire complex plane. The reason we see white points when using the above algorithm is that it takes some points longer than 25 or 50 iterations to escape.



FIG. 1a. The Julia set of $.3e^z$.

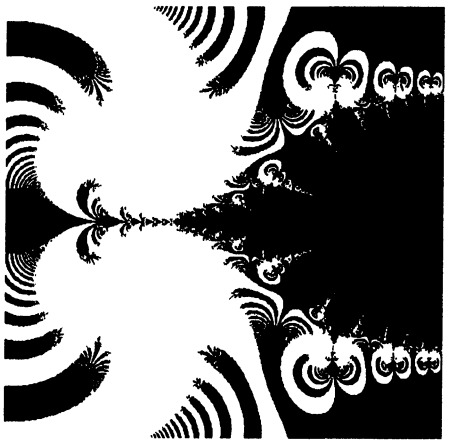


FIG. 1b. Julia set of $.4e^z$.

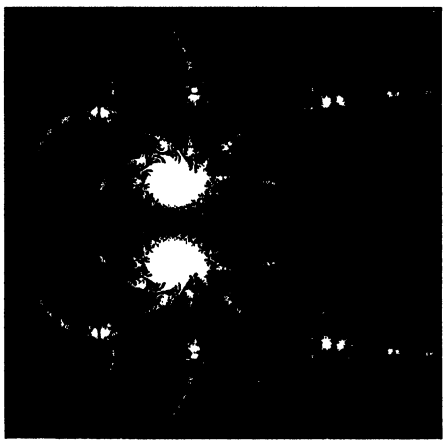


FIG. 1c. Fig. 1b. recomputed with $N = 50$.

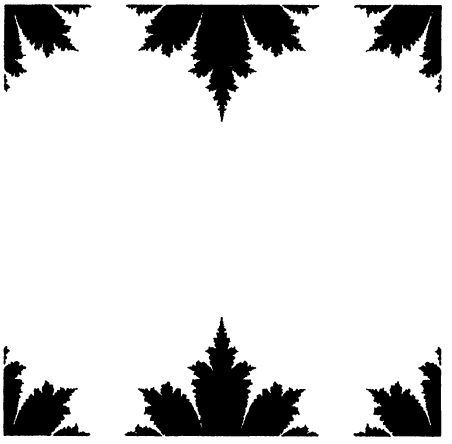


FIG. 2a. The Julia set of $.66i \cos z$.

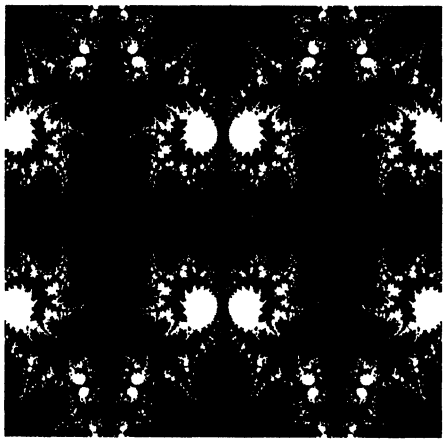


FIG. 2b. The Julia set of $.68i \cos z$.

Explosions occur in other families of transcendental functions too. For example, the Julia sets of the complex sine or cosine functions may be computed using the above algorithm with one slight modification. For sine and cosine the Julia set may also be defined as the closure of the points whose orbits tend to infinity with the imaginary part tending to $\pm\infty$. Hence we color a point black if its orbit enters the region $|\text{Im } z| > 50$. FIGURE 2 shows the result of applying this algorithm to $.66i \cos(z)$ and $.68i \cos(z)$. FIGURE 3 displays the Julia set of $\sin(z)$ and $(1 + .1i)\sin(z)$. In both cases we note that the Julia sets undergo dramatic changes. These are the explosions which we describe in this paper.

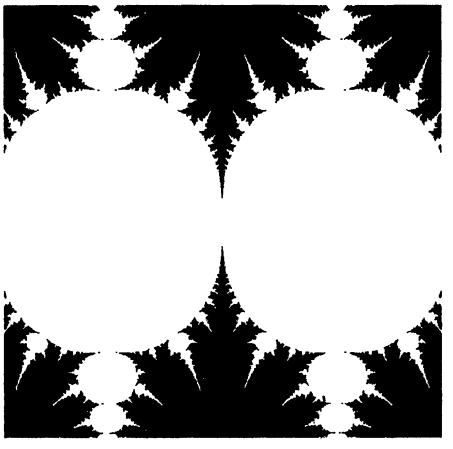


FIG. 3a. The Julia set of $\sin z$.

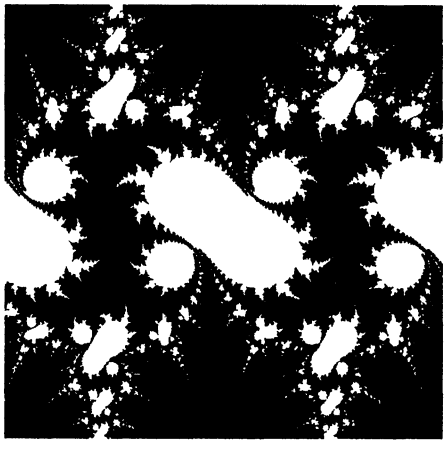


FIG. 3b. The Julia set of $(1 + .1i)\sin z$.

2. Preliminaries. Before looking at the specific case of λe^z, let us first consider some basic definitions from dynamical system theory. First, a *fixed point* for a function F, as the name implies, is a point z such that $F(z) = z$. The point z is *periodic* if there exists an integer n such that $F^n(z) = z$. We define the smallest positive such n as the *period* of z and note immediately that any fixed point, z, is periodic with period one, and, in fact, $F^n(z) = z$ for all n.

Of particular interest to us as we consider Julia sets will be the behavior of orbits nearby certain periodic points. We first distinguish several different types of periodic points. First, if z_0 is a periodic point of period n, and $0 < |(F^n)'(z_0)| < 1$, then z_0 is said to be an *attracting periodic point*. In the special case when $|(F^n)'(z_0)| = 0$, we call z_0 a *super attracting periodic point*. If z_0 is a periodic point of period n and $|(F^n)'(z_0)| > 1$, then z_0 is defined to be a *repelling periodic point*. The intermediate case, $|(F^n)'(z_0)| = 1$, gives *indifferent periodic points*. As we would expect, points nearby an attracting fixed point tend, upon iteration, towards that point. In the case of a repeller, points nearby the fixed point eventually escape from that point upon iteration.

More specifically, we state

PROPOSITION 2.1. *If z_0 is an attracting fixed point, then there is an open disk U about z_0 such that if z is an element of U, then $\lim_{n\to\infty}(F^n)(z) = z_0$.*

Proof. Since z_0 is an attracting fixed point, there exists $\delta > 0$ and $\mu < 1$ such that if $|z - z_0| < \delta$, then $|F'(z)| < \mu$. Hence it follows that if $z_1, z_2 \in B_\delta(z_0)$, the ball of radius δ about z_0, then

$$|F(z_1) - F(z_2)| < \mu|z_1 - z_2| < |z_1 - z_2|.$$

Therefore, $F(B_\delta(z_0)) \subset B_{\mu\delta}(z_0)$. By the Contraction Mapping Principle, F is a contraction on $B_\delta(z_0)$, and so all points in the ball tend to the fixed point z_0 under iteration.

In the above proof, the set $D = B_\delta(z_0) - F(B_\delta(z_0))$ is called a *fundamental domain*. This means that, if $z \in B_\delta(z_0)$, $z \neq z_0$, then there is a unique $w \in D$ and $n \geq 0$ such that $F^n(w) = z$. This means that each forward orbit in $B_\delta(z_0)$ which is not the fixed point passes through D exactly once.

To see why this is true, we note that, as above,

$$F^n(B_\delta(z_0)) \subset B_{\mu^n\delta}(z_0).$$

Hence there is a smallest integer $j \geq 0$ such that $z \in F^j(B_\delta(z_0))$ but $z \notin F^{j+1}(B_\delta(z_0))$. Then we may let $w \in D$ be such that $F^j(w) = z$ and $n = j$. See FIGURE 4.

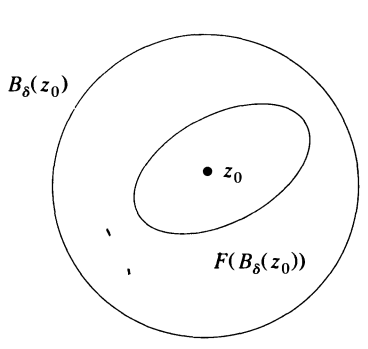


FIG. 4. $B_\delta(z_0) - F(B_\delta(z_0))$ is a fundamental domain.

We call the set of all points whose orbits tend to a given fixed point the *basin of attraction* of that point. Clearly, the basin of attraction is an open set.

In the case of a repelling fixed point, nearby points eventually escape from the repeller upon iteration. In contrast to the previous proposition, we state

PROPOSITION 2.2. *If z_0 is a repelling fixed point, then there exists a disk U about z_0 such that if $z \in U (z \neq z_0)$, then there exists $k > 0$ such that $F^k(z) \notin U$.*

Proof. Since z_0 is a repelling fixed point, we have $|F'(z_0)| > 1$. Hence, by the Inverse Function Theorem, there exists a neighborhood of z_0 on which a branch of F^{-1} exists. We compute

$$\left|(F^{-1})'(z_0)\right| = 1/\left|F'\left(F^{-1}(z_0)\right)\right| = 1/\left|F'(z_0)\right| < 1.$$

We conclude that z_0 is an attracting fixed point for F^{-1} and hence, by our previous proposition, there exists $\delta > 0$ such that if $z \in B_\delta(z_0)$ then $F^{-1}(z) \in B_\delta(z_0)$ and $(F^{-1})^n(z) \to z_0$.

Now assume that for any neighborhood U of z_0, there exists z^*, $z^* \neq z_0$, which satisfies $F^k(z^*) \in U$ for all $k \geq 0$. We may assume that $U = F^{-1}(B_\delta(z_0))$ for δ sufficiently small. As above, $B_\delta(z_0) - F^{-1}(B_\delta(z_0))$ is a fundamental domain for F^{-1}. Hence there exists $w \in B_\delta(z_0) - F^{-1}(B_\delta(z_0))$ and $n > 0$ such that $(F^{-1})^n(w) = z^*$. But then $F^n(z^*) = w \notin U$. This contradiction establishes the result. q.e.d.

In order to give a more precise definition of the Julia set, we now shift our attention to some important concepts from traditional complex analysis. Let $\mathcal{F}$ be a family of analytic functions defined on an open set $U \subset \mathbf{C}$. For our purposes, $\mathcal{F}$ will usually be the family of iterates of a given complex function. The family $\mathcal{F}$ is said to be *normal* on U if every sequence in $\mathcal{F}$ has a subsequence that either (1) converges uniformly on every compact subset of U, or (2) converges uniformly to ∞ on U.

Of particular interest to us later on will be

THEOREM 2.3 (Montel). *If* $\bigcup_{f \in \mathcal{F}}[f(U)]$ *omits two or more points in the complex plane, then* $\mathcal{F}$ *is a normal family on* U.

Equivalently, if $\mathcal{F}$ is not normal on U, then

$$\bigcup_{f \in \mathcal{F}} [f(U)]$$

omits at most one point in the complex plane.

In dealing with entire, transcendental functions like λe^z we have access to a particularly powerful tool. Recall that such functions have an essential singularity at ∞. Hence, they satisfy the hypothesis for the Great Picard Theorem.

THEOREM 2.4. *Suppose an analytic function* F *has an essential singularity at* $z = a$. *Then for any neighborhood* U *of* a *and for all* $z^* \in \mathbf{C}$ (*with at most one exception*) *there exist infinitely many* $z \in U$ *such that* $F(z) = z^*$.

We are now ready to give a formal definition of the Julia set. Let us consider $\mathcal{F} = \{F^n\}$ where F^n is the nth iterate of F. We define the Julia set of F, $J(F)$, by

$$J(F) = \{z | \{F^n\} \text{ is not normal on any neighborhood of } z\}.$$

We make the observation here that the Julia set is completely invariant. That is, if a point is contained in $J(F)$, then so are all of its images and all of its preimages. In addition, from this definition we may immediately conclude that all repelling periodic points are contained in $J(F)$. In fact, we can show

PROPOSITION 2.5. *Closure*{*repelling periodic points*} $\subset J(F)$.

Proof. Suppose z_0 belongs to the closure of the repelling periodic points. If U is an open neighborhood of z_0, then there exists $z^* \in U$ such that z^* is a repelling periodic point. Without loss of generality we will assume that z^* is a repelling fixed point. Hence, there exists a neighborhood V of z^* such that for $z \in V$ $(z \neq z^*)$ there exists $k > 0$ such that $F^k(z) \notin V$. At the same time, we

know that $F^k(z^*) = z^* \in V$ for all values of k. Hence, $\{F^n\}$ cannot converge uniformly on V. It follows that $\{F^n\}$ is not normal on U and, therefore, $z_0 \in J(F)$.

In contrast to the above proposition, the attracting fixed points and their basins of attraction are never contained in the Julia set. This follows easily from our previous work when we note that any point in the basin of attraction of a fixed point z_0 has about it a neighborhood U such that if $z \in U$, then $\lim_{n\to\infty} F^n(z) = z_0$. Hence, F^n converges uniformly to the constant function $g(z) = z_0$ on U. We conclude that F^n is normal at z_0 and, therefore, $z_0 \notin J(F)$. Similar arguments give the same result for periodic points.

We conclude this section by discussing some of the elementary dynamical properties of the exponential function $E_\lambda(z) = \lambda e^z$ with $\lambda > 0$. As a dynamical system on the real line, this map has two quite distinct dynamical behaviors, depending on whether $0 < \lambda < 1/e$ or $\lambda > 1/e$. This can be seen immediately by looking at the graphs of E_λ (see FIGURE 5). When $\lambda < 1/e$, E_λ has a pair of fixed points, q_λ and p_λ, with $0 < q_\lambda < p_\lambda$. Note that q_λ is an attracting fixed point since $E'_\lambda(q_\lambda) < 1$. On the other hand, p_λ is a repelling fixed point since $E'_\lambda(p_\lambda) > 1$. If $x > p_\lambda$, then $E_\lambda(x) > x$ and it follows that $E^n_\lambda(x) \to \infty$ as $n \to \infty$. If, however, $x < p_\lambda$, $x \neq q_\lambda$, then

$$|E_\lambda(x) - q_\lambda| < |x - q_\lambda| \qquad (*)$$

so $E^n_\lambda(x) \to q_\lambda$ as $n \to \infty$. When $\lambda > 1/e$, the dynamics of E_λ are quite different. For all $x \in \mathbf{R}$, we have $E_\lambda(x) > x$, so that $E^n_\lambda(x) \to \infty$ as $n \to \infty$. Thus all points have orbits which tend to ∞ when $\lambda > 1/e$.

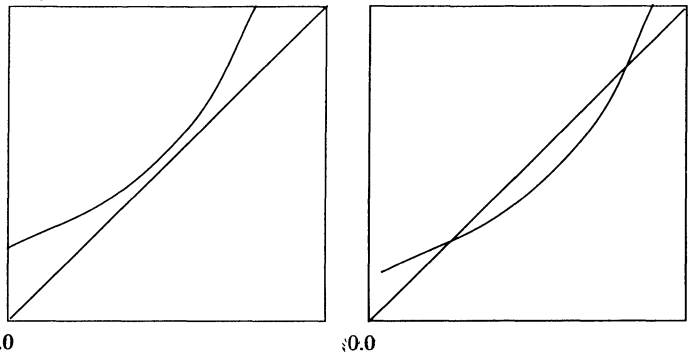


FIG. 5. The graphs of $\lambda \exp z$ for $\lambda > 1/e$ and $\lambda < 1/e$.

The change in the dynamics of E_λ which occurs at $\lambda = 1/e$ is an example of a *bifurcation*. In this case it is a relatively simple bifurcation known as a saddle-node bifurcation. In this paper we will investigate the effects of this bifurcation in the complex plane. We will show that the global ramifications of this bifurcation are quite spectacular in $\mathbf{C}$: the Julia set of E_λ explodes as λ increases through $1/e$.

Before describing the Julia set of E_λ, we recall some of the basic mapping properties of the exponential function:

1. $E_\lambda(x + iy) = \lambda e^x e^{iy} = \lambda e^x(\cos y + i \sin y)$ (Euler's formula.)
2. Consequently, E_λ maps vertical lines $x = c$ to circles of radius λe^c centered at the origin, and horizontal lines $y = c$ to rays $\theta = c$ emanating from the

origin. In particular, any rectangle with sides parallel to the axes and vertical length 2π is mapped onto an annular region surrounding the origin.

3. E_λ is $2\pi i$-periodic.

3. The Tame Exponential. In this section we will describe the Julia set of E_λ when $\lambda < 1/e$. We will see that $J(E_\lambda)$ is quite small in this case: most points have orbits which are stable when $\lambda < 1/e$. Recall that when $\lambda < 1/e$, E_λ has two fixed points on $\mathbf{R}$: an attracting fixed point at q_λ and a repelling fixed point at p_λ. Since $E_\lambda(1) = \lambda e$ and $E_\lambda(-\log \lambda) = 1$, we have

$$0 < q_\lambda < 1 < -\log \lambda < p_\lambda. \qquad (*)$$

Consider the half plane

$$H = \{z | \operatorname{Re} z < -\log \lambda\}.$$

If $z \in H$, then

$$|E'_\lambda(z)| = |E_\lambda(z)| = \lambda \exp(\operatorname{Re} z) < 1.$$

Hence E_λ contracts H into the unit disk, which, since $-\log \lambda > 1$, is completely contained in H. Thus, by the Contraction Mapping Theorem, all points in H tend under iteration to a fixed point, which must be q_λ. Let us denote the basin of attraction of q_λ by $W(q_\lambda)$. Thus, $H \subset W(q_\lambda)$. Clearly, all points in $W(q_\lambda)$ have stable orbits since they all tend asymptotically to the same point. As a consequence, we will be mainly interested in the complement of $W(q_\lambda)$. As we shall see, this set is the Julia set, $J(E_\lambda)$.

Let us investigate the dynamics of E_λ in the complement of $W(q_\lambda)$. Using (*), we may choose ν such that

$$1 < E_\lambda(\nu) < -\log \lambda < \nu < p_\lambda.$$

Let H_ν denote the half plane $\operatorname{Re} z < \nu$. Arguing as before, E_λ maps H_ν inside H and so H_ν is also contained in $W(q_\lambda)$. Note that if $\operatorname{Re} z > \nu$, then

$$|E'_\lambda(z)| = |E_\lambda(z)| > \lambda e^\nu > 1.$$

This fact is important since it shows that E_λ is an expanding map on the complement of H_ν.

Before discussing the properties of the complement of $W(q_\lambda)$ we first show that, for $\lambda < 1/e$, this set is quite small. To do this we need an important lemma that we will use several times in the sequel. (We note that this result holds for all λ, not just $\lambda < 1/e$.) Let $B_\delta(z_0)$ denote the open ball of radius δ centered at z_0.

LEMMA 3.1 (Expansion Lemma). *Suppose* $|E'_\lambda(z)| > \mu$ *for all* $z \in B_\delta(z_0)$ *where* $\delta < \pi$. *Then there is an open set* $U \subset B_\delta(z_0)$ *such that* $E_\lambda\colon U \to B_{\mu\delta}(E_\lambda(z_0))$ *is a homeomorphism.*

Proof. Since $\delta < \pi$, E_λ is one-to-one on $B_\delta(z_0)$. Hence we may define the inverse map $L\colon E_\lambda(B_\delta(z_0)) \to B_\delta(z_0)$. From the chain rule it follows that $|L'(z)| < 1/\mu$ for all $z \in E_\lambda(B_\delta(z_0))$. Thus, if $|z_0 - z_1| = \delta$, we have

$$\begin{aligned} \delta = |z_0 - z_1| &= |L(E_\lambda(z_0)) - L(E_\lambda(z_1))| \\ &\le \frac{1}{\mu} |E_\lambda(z_0) - E_\lambda(z_1)| \end{aligned}$$

It follows that E_λ maps $|z_0 - z_1| = \delta$ to a curve which contains $B_{\mu\delta}(E_\lambda(z_0))$ in its interior. By the Maximum Principle, $E_\lambda(B_\delta(z_0)) \supset B_{\mu\delta}(E_\lambda(z_0))$.

Thus we see that, as long as the orbit of a point remains in the complement of H_ν, where $|E_\lambda'(z)| = |E_\lambda(z)| \geq \nu > 1$, successive iterations of E_λ tend to expand neighborhoods of the original point. We may apply this idea repeatedly to prove the following result.

THEOREM 3.2. *Suppose z_0 lies in the complement of $W(q_\lambda)$ and suppose U is an open set containing z_0. Then $U \cap W(q_\lambda) \neq \varnothing$.*

Proof. Assume that $E_\lambda^n(U) \cap W(q_\lambda) = \varnothing$ for all n. It follows that $E_\lambda^n(U) \cap H_\nu = \varnothing$ for all n. Hence we have

$$|E_\lambda'(z)| > \mu > 1$$

for all $z \in U$, where $\mu = \lambda e^\nu$. Choose δ so that $B_\delta(z_0) \subset U$. Then the Expansion Lemma implies that $E_\lambda(B_\delta(z_0)) \supset B_{\mu\delta}(E_\lambda(z_0))$. Now $B_{\mu\delta}(E_\lambda(z_0))$ does not meet H_ν, for otherwise we have a contradiction. It follows that we may apply the Expansion Lemma again to this disk. Continuing in this fashion, and using the fact that $E_\lambda^n(z_0) \notin H_\nu$, it follows that we may find a disk of radius $\mu^n\delta$ about $E_\lambda^n(z_0)$ which does not meet H_ν. If we choose n large enough so that $\mu^n\delta > 2\pi$, then this disk must meet one of the horizontal lines of the form $y = (2k + 1)\pi$, where $k \in \mathbf{Z}$. But these lines are mapped by E_λ onto the negative real axis, which lies in H_ν. This contradiction establishes the result. q.e.d.

COROLLARY 3.3. *$W(q_\lambda)$ is open and dense in* $\mathbf{C}$.

As a consequence, the complement of $W(q_\lambda)$ is a nowhere dense subset of the right half plane. We use this fact to prove:

PROPOSITION 3.4. *The complement of $W(q_\lambda)$ is precisely $J(E_\lambda)$.*

Proof. We show first that the complement of $W(q_\lambda) \subset J(E_\lambda)$. Suppose z_0 belongs to the complement of $W(q_\lambda)$. Since the complement of $W(q_\lambda)$ is nowhere dense we know that for each neighborhood U of z_0 there exists $z \in U$ such that $z \in W(q_\lambda)$. Now, with n sufficiently large, we can make $|E_\lambda^n(z) - q_\lambda|$ as small as we wish. In particular, we know that there exists $N > 0$ such that $E_\lambda^n(z)$ lies to the left of $-\log \lambda$ for all $n > N$. At the same time, notice that $E_\lambda^n(z_0)$ must remain to the right of $-\log \lambda$ for all n. Therefore, the family of functions $\{E_\lambda^n\}$ is not normal at z_0. We conclude that the complement of $W(q_\lambda)$ is contained in $J(E_\lambda)$.

For the opposite direction, let z_0 be in $J(E_\lambda)$. By definition, $\{E_\lambda^n\}$ is not normal on any neighborhood of z_0. Let us suppose $z_0 \in W(q_\lambda)$. Then there exists a neighborhood, U, of z_0 such that $E_\lambda^n(U) \to q_\lambda$ since $W(q_\lambda)$ is open. Hence, $\{E_\lambda^n\}$ converges to the constant function $g(z) = q_\lambda$ on U. Thus, $\{E_\lambda^n\}$ is normal on U. This contradiction shows that z_0 is contained in the complement of $W(q_\lambda)$.

Note that, as a result of this proposition, all points which escape to ∞ under iteration of E_λ necessarily lie in $J(E_\lambda)$. There are other points, like p_λ, which do not escape, but which nevertheless lie in the Julia set. From the graph of E_λ (FIG. 5), we see that in the case of p_λ itself, there are points arbitrarily nearby which do escape. This is a general phenomenon: the Julia set may also be described as the closure of the set of points which escape to ∞ under iteration of E_λ. We stress that this is not the case for *all* functions. In particular, for polynomials, the escaping points are never in the Julia set since ∞ acts as an attracting fixed point in $\mathbf{C} \cup \{\infty\}$. For entire functions, however, the essential singularity at ∞ makes the situation quite different: Picard's Theorem implies that neighborhoods of ∞ are mapped over all of $\mathbf{C} - \{0\}$ by E_λ. This suggests that escaping points are in the Julia set in this case.

Let $\mathrm{Esc}(\lambda)$ be the closure of the set of points which escape to ∞ under iteration of E_λ. We now prove

THEOREM 3.5. *If $\lambda < 1/e$, then $J(E_\lambda) = Esc(\lambda)$.*

Proof. Clearly, if $z_0 \in \mathrm{Esc}(\lambda)$, then $z_0 \notin W(q_\lambda)$ so $z_0 \in J(E_\lambda)$. Hence, $\mathrm{Esc}(\lambda) \subset J(E_\lambda)$. For the other direction, suppose $z_0 \in J(E_\lambda)$. If U is any neighborhood of z_0 we must produce a $z \in U$ such that $|E^n(z)| \to \infty$. Since $\{E_\lambda^n\}$ is not normal on any neighborhood of z_0, Montel's Theorem tells us that $\bigcup_n [E_\lambda^n(U)]$ omits at most one point in $\mathbf{C}$. In particular, there is a point $z^* \in \mathbf{R}$ with $z^* > p_\lambda$ such that $z^* \in \bigcup_n [E_\lambda^n(U)]$. Thus, there exists $z \in U$ such that $E_\lambda^j(z) = z^*$ for some j. Consequently, $z \in \mathrm{Esc}\ \lambda$. Hence $z_0 \in \mathrm{Esc}(\lambda)$ and $J(E_\lambda) \subset \mathrm{Esc}(\lambda)$.

We now describe the geometry of the Julia set of E_λ when $\lambda < 1/e$. Certainly, $J(E_\lambda)$ contains the interval $[p_\lambda, \infty)$, since p_λ is a repelling fixed point and all other points in the interval tend to ∞ under iteration. By $2\pi i$-periodicity and invariance of $J(E_\lambda)$, the Julia set also contains the $2\pi i$ translates of this interval as well as all preimages of these lines under E_λ^n. These are by no means all of the points in $J(E_\lambda)$; this set contains what we call a "Cantor bouquet." We will sketch the geometric construction of this set. For more details, we refer the reader to [**DT**].

Let C_0 denote the half-plane $\mathrm{Re}\ z \geq \nu$. C_0 is the complement of H_ν. Let us remove from C_0 all points which are mapped into H_ν by E_λ. The regions that remain are therefore defined by applying various branches of the inverse transformation of E_λ given by

$$L_{\lambda,k}(z) = \log_k(z/\lambda)$$

to C_0. Here $\log_k$ is the branch of the logarithm function defined on $\mathbf{C} - \{x \in \mathbf{R} | x \leq 0\}$ and taking values in the strip $(2k-1)\pi < \mathrm{Im}\ z < (2k+1)\pi$. Each $L_{\lambda,k}$ maps C_0 onto a region shaped like a "tongue" as depicted in FIGURE 6. Let us denote by T_k the image of C_0 under $L_{\lambda,k}$. Let $C_1 = \bigcup T_k$. C_1, therefore, consists of all points in C_0 which remain in C_0 for at least one iteration of E_λ. We think of C_1 as being obtained from C_0 by the removal of an open set which consists of "fingers" extending to the right and separating the T_k's. Notice that these fingers are mapped by E_λ into $H_\nu \subset W(q_\lambda)$.

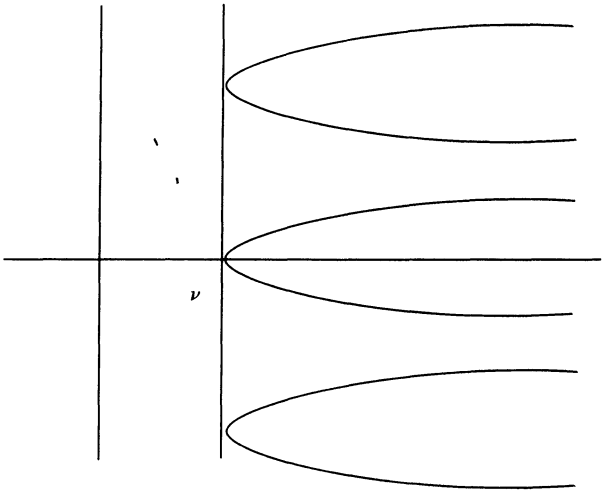


FIG. 6. The tongues in C_0.

We may obtain $J(E_\lambda)$ by iterating this procedure. Apply each $L_{\lambda,k}$ to C_1. The result is an infinite collection of tongues which lie in T_k. Equivalently, E_λ expands T_k over C_0 in a one-to-one fashion and so there are regions in T_k which are mapped onto each T_j. Let us denote the image of C_1 under all of the $L_{\lambda,k}$'s by C_2. We again think of C_2 as being obtained from C_1 by the removal of a collection of fingers in C_1 which extend to the right. See FIGURE 7. C_2 consists of all points in C_0 whose first two E_λ-images lie in C_0.

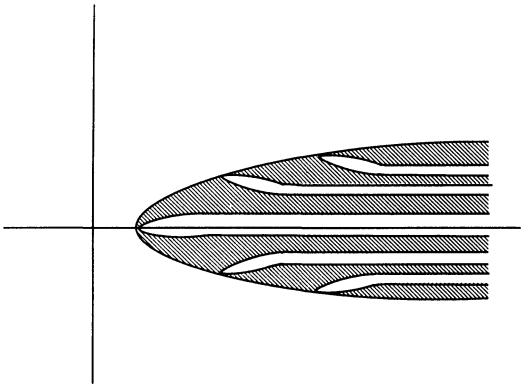


FIG. 7. The construction of C_2.

Continuing in this fashion, we denote by C_n the images of C_{n-1} under all of the $L_{\lambda,k}$. The reader will undoubtedly recognize this construction as similar to that of the construction of the Cantor Middle-Thirds set as encountered in elementary analysis. Instead of removing open intervals from the middle of closed intervals, we are removing open sets which separate C_{n-1} into an infinite collection of tongues which we denote by C_n. Clearly, the C_n are nested closed sets. Note that

$$J(E_\lambda) = \bigcap_{n=0}^{\infty} C_n.$$

This follows since any point which does not lie in some C_n is mapped into H_ν under E_λ^n.

It can be shown [**DK**] that this intersection consists of a collection of curves in **C**. Each curve, like the interval $[p\lambda, \infty)$, is homeomorphic to the non-negative real axis. The collection of these curves forms a perfect set, and hence our terminology Cantor bouquet.

Remark. It is not true that all of the nested sequences of tongues described above actually yield a non-empty intersection; some of the tongues nest down to ∞ as $n \to \infty$. The exact situation is described in [DK]. For our purposes, the important thing is the dramatic difference between the structure of this set and its counterpart for $\lambda > 1/e$ described in the next section.

4. The Chaotic Exponential. In this section we will complete the proofs that $J(E_\lambda)$ is the closure of the set of repelling periodic points as well as the closure of the set of escaping points when $\lambda > 1/e$. Thus far we have shown that

$$J(E_\lambda) = \{z | \{E_\lambda^n\} \text{ is not normal at } z\} \supset \overline{\{\text{repelling periodic points}\}}.$$

In addition, for $\lambda < 1/e$, we have seen that $J(E_\lambda) = \text{Esc}(\lambda)$. As before, the

essential singularity at ∞ suggests that we can adopt this third characterization of the Julia set when $\lambda > 1/e$ as well.

Recall that the exponential is an expanding map in the half plane $\operatorname{Re} z > -\log \lambda$ in the sense that $|E_\lambda'(z)| = \lambda \exp(\operatorname{Re} z) > 1$. Geometrically, this means that E_λ expands areas in this region. This is the basic fact which we exploit throughout this section.

If $\operatorname{Re}(E_\lambda(z_0)) > \log(3/\lambda)$, then $|E_\lambda'(z_0)| > 3$. It follows from the Expansion Lemma that E_λ expands small neighborhoods of z_0 by a factor of 3. We may apply this idea repeatedly to prove:

PROPOSITION 4.1. *Suppose*

$$\operatorname{Re}\left(E_\lambda^j(z_0)\right) > \log(3/\lambda) + 3\pi$$

for $j = 0, 1, 2, \ldots$. Let U be any open set containing z_0 and let $S_{2\pi}(E_\lambda^n(z_0))$ be the square with center $E_\lambda^n(z_0)$ and side of length 2π. Then there exists $N > 0$ such that, if $n > N$, there exists an open set $U_n \subset U$ for which $E_\lambda^n\colon U_n \to S_{2\pi}(E_\lambda^n(z_0))$ is a homeomorphism.

Proof. Since $\operatorname{Re}(E_\lambda^j(z_0)) > 3\pi + \log(3/\lambda)$, it follows that, if $z \in B_{3\pi}(E_\lambda^j(z_0))$, then

$$|E_\lambda'(z)| = \lambda \exp(\operatorname{Re} E_\lambda(z)) > 3.$$

Now suppose $B_\delta(z_0) \subset U$. Choose N so that $3^{N-1}\delta < \pi$ but $3^N\delta \geq \pi$. Apply the Expansion Lemma repeatedly to find an open set $W_N \subset B_\delta(z_0)$ that is mapped homeomorphically onto $B_{3^N\delta}(E_\lambda^N(z_0))$ by E_λ^N. Now $B_{3^N\delta}(E_\lambda^N(z_0)) \supset B_\pi(E_\lambda^N(z_0))$. Again apply the Expansion Lemma. We find, for each $k > 0$, an open set $V_k \subset B_\pi(E_\lambda^N(z_0))$ such that $E_\lambda^k\colon V_k \to B_{3\pi}(E_\lambda^{N+k}(z_0))$ is a homeomorphism.

Since $B_{3\pi}(E_\lambda^{N+k}(z_0)) \supset S_{2\pi}(E_\lambda^{N+k}(z_0))$, we may restrict E_λ^k to $V_k' \subset V_k$ so that $E_\lambda^k\colon V_k' \to S_{2\pi}(E_\lambda^{N+k}(z_0))$ is also a homeomorphism. Then, if $n = N + k$, set $U_n = E_\lambda^{-N}(V_k')$. It follows that $E_\lambda^n\colon U_n \to S_{2\pi}(E_\lambda^n(z_0))$ is a homeomorphism. q.e.d.

Now suppose z_0 is an escaping point. Since $|E_\lambda^n(z_0)| \to \infty$, it follows that $\operatorname{Re} E_\lambda^n(z_0) \to \infty$ as well. Thus all of the escaping orbits tend to ∞ in the "same" direction—to the right.

LEMMA 4.2. *Let $z_0 \in \operatorname{Esc}(\lambda)$ and suppose U is an open set containing z_0. Then there exists $z \in U$ and $n \geq 0$ such that $E_\lambda^n(z) \in \mathbf{R}$.*

Proof. Since z_0 escapes, we may assume that $\operatorname{Re} E_\lambda^j(z_0) > \log(3/\lambda) + 3\pi$ for all $j \geq 0$. Hence the previous Proposition applies and there exists an open set $U_n \subset U$ such that $E_\lambda^n\colon U_n \to S_{2\pi}(E_\lambda^n(z_0))$ is a homeomorphism. But $S_{2\pi}(E_\lambda^n(z_0))$ must meet some horizontal line of the form $y = k\pi$, $k \in \mathbf{Z}$. Then $E_\lambda^{n+1}(U_n) \cap \mathbf{R} \neq \phi$.

Now we can prove

THEOREM 4.3. *If $\lambda > 1/e$, then $J(E_\lambda) = \operatorname{Esc}(\lambda)$.*

Proof. The proof that $J(E_\lambda) \subset \operatorname{Esc}(\lambda)$ is exactly the same as in the case $\lambda < 1/e$. Hence we only prove the converse. Let $z_0 \in \operatorname{Esc}(\lambda)$. If U is an open set containing z_0, then we must show $\{E_\lambda^n\}$ is not normal on U. By the previous lemma and the fact that all real points escape when $\lambda > 1/e$, we may assume that $z_0 \in \mathbf{R}$ and $z_0 > \log(3/\lambda) + 3\pi$.

For simplicity, we write $z_n = E_\lambda^n(z_0)$. Note that $z_{n+1} > z_n$. By Proposition 4.1 we may choose $n > 0$ so that $E_\lambda^n(U) \supset S_{2\pi}(z_n)$. The mapping properties of E_λ tell

us that $E_\lambda(S_{2\pi}(z_n))$ is the annulus given by

$$e^{-\pi} z_{n+1} \leq |z| \leq e^{\pi} z_{n+1}.$$

Since $e^{-\pi} < 1/10$ and $e^{\pi} > 10$, it follows that the annulus $E_\lambda(S_{2\pi}(z_n))$ contains $S_{2\pi}(z_{n+1})$. Continuing in this fashion, we find that $E_\lambda^k(S_{2\pi}(z_n))$ contains the annulus

$$e^{-\pi} z_{n+k} \leq |z| e^{\pi} z_{n+k}.$$

We denote this annulus by A_k; since $z_{n+k} \to \infty$ as $k \to \infty$, it follows that the A_k grow very large with k. Thus, as we let k increase we can produce points in U which get mapped near the left outer radius of A_k as well as those that get mapped near the right outer radius. These points, in turn, get mapped "very close to" 0 and "very close to" ∞. We conclude that $\{E_\lambda^n\}$ is not normal on U. q.e.d.

With the above proof we have two equivalent definitions of the Julia set of E_λ, $\lambda > 0$. We now add a third. Recall that in our preliminary section we showed that the closure of the set of repelling periodic points was contained in $J(F)$ for any complex function F. We now show inclusion the other way for our function E_λ. We first state a variation of the Schwarz Lemma for use in our proof.

LEMMA 4.4 (Schwarz). *Let U be open, bounded and simply connected in* $\mathbf{C}$. *If* $F\colon \mathbf{C} \to \mathbf{C}$ *is analytic and if* $F(\overline{U}) \subset U$, *then there exists a fixed point* $z_0 \in U$, *and* $|F'(z_0)| < 1$.

We are now ready to prove

PROPOSITION 4.5. $J(E_\lambda) \subset$ *closure*{*repelling periodic points*}.

Proof. Let $z_0 \in J(E_\lambda)$ and suppose W is an arbitrary neighborhood of z_0. We will show that there is a repelling periodic point in W. Since p_λ is a repelling fixed point for E_λ, the Inverse Function Theorem implies that there exists a neighborhood U_0 of p_λ on which E_λ^{-1} exists. In fact, we can find U_0 such that $U_1 = E_\lambda^{-1}(U_0) \subset U_0$. Continued iteration produces a nested sequence of neighborhoods of p_λ, $U_0 \supset U_1 \supset U_2 \supset \cdots$.

Montel's Theorem implies that since $\{E_\lambda\}$ is not normal at z_0, then $\bigcup_n [E_\lambda^n(W)]$ omits at most one point in $\mathbf{C}$. Since $\bigcup_n [E_\lambda^n(W)]$ already omits zero ($E_\lambda^n(z) \neq 0$ for all $z \in \mathbf{C}$), then there exists $z_1 \in W$ and $m > 0$ such that $E_\lambda^m(z_1) = p_\lambda$. Now since p_λ is a repelling fixed point, $\{E_\lambda^n\}$ is not normal on any neighborhood of p_λ and, by the same reasoning as above, there exists $z_2 \in U_0$ and $n > 0$ such that $E_\lambda^n(z_2) = z_1$. Therefore, $E_\lambda^{m+n}(z_2) = p_\lambda$.

We know $E_\lambda'(z) = E_\lambda(z) \neq 0$ for all $z \in \mathbf{C}$. Combining this fact with use of the chain rule, we conclude $(E_\lambda^m)'(z_1) \neq 0$ and $(E_\lambda^n)'(z_2) \neq 0$ for any $m, n \in \mathbf{Z}$. We apply the Inverse Function Theorem again to claim that there exists a neighborhood V of z_2 which is mapped in a one-to-one fashion onto a neighborhood of z_1 in W. Hence, $E_\lambda^n(V) \subset W$. Since $(E_\lambda^m)'(z_1) \neq 0$, we may use similar arguments (choosing V smaller if necessary) to show that E_λ^{m+n} maps V diffeomorphically onto U_j for some j. Further, $E_\lambda^{m+n+j}(V) = U_0$ since $U_j = (E_\lambda^{-1})^j(U_0)$. Now $E_\lambda^{n+m+j}\colon V \to U_0$ is one-to-one and onto with local inverse $(E_\lambda^{m+n+j})^{-1}$. Using the Schwarz Lemma we conclude $(E_\lambda^{m+n+j})^{-1}$ has an attracting fixed point in V. See FIGURE 8. Equivalently, E_λ^{m+n+j} has a repelling fixed point in V which implies E_λ has a repelling periodic point, z^*, in V. But, E_λ^n maps V diffeomorphically onto a neighborhood of $z_1 \in W$. Therefore, part of the orbit of z^* passes through W. Now, since any point in a repelling periodic orbit is itself a repelling periodic point, we conclude that there is a repelling periodic point in W.

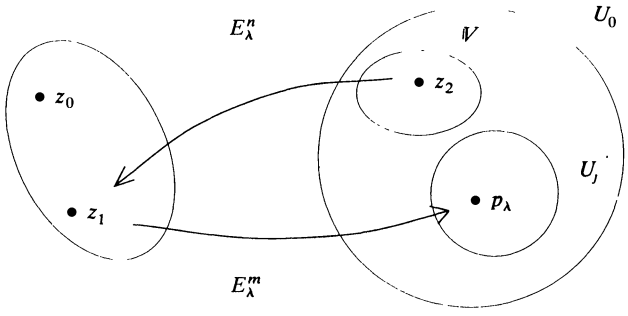


FIG. 8

We comment here that although our proof specifically deals with $E_\lambda(z)$, this argument may be adapted to show that $J(F)$ is the closure of the set of repelling periodic points for any function with at least one repelling periodic orbit.

What does the Julia set of E_λ look like when $\lambda > 1/e$? Recall from section 2 that when $\lambda > 1/e$, the function $E_\lambda(x)$, $x \in \mathbf{R}$, has no fixed points. In fact, $E_\lambda^n(x) \to \infty$ for all values of x. Hence, in immediate contrast to the $\lambda > 1/e$ case, we have $\mathbf{R} \subset J(E_\lambda)$. We can show that there are many other points in $J(E_\lambda)$ as well.

Let $\mathscr{L}$ be a vertical line segment of length 2π in the left half plane (i.e., far removed from the Cantor bouquet described in the previous section). Recall that E_λ maps this segment to a circle about the origin. This circle necessarily crosses the real axis twice. Hence, there are two points which now, upon further application of E_λ, tend to ∞. If we call these two crossing points x_1 and x_2, we see that there must be two points, z_1 and z_2, in $\mathscr{L}$ such that $E_\lambda(z_1) = x_1$ and $E_\lambda(z_2) = x_2$. We now have $E_\lambda^n(z_1) \to \infty$ and $E_\lambda^n(z_2) \to \infty$ as $n \to \infty$. We continue choosing arbitrary line segments in this manner to produce pairs of points in the Julia set throughout the complex plane. In fact, we see that all horizontal lines of the form $\operatorname{Im} z = n\pi$ where $n \in \mathbf{Z}$ are contained in $J(E_\lambda)$.

Notice that we cannot use this argument when $\lambda < 1/e$ since in that case both z_1 and z_2 would be attracted to the fixed point q_λ instead of escaping to ∞. This roadblock is lifted precisely when λ values increase through $1/e$—hence, the explosion.

With the above reasoning, we have shown that $J(E_\lambda)$ for $\lambda > 1/e$ admits many more points than in the $\lambda < 1/e$ case. In fact, for these higher λ values, the Julia set is the whole plane!

Fatou first conjectured this in the early 1900's and in 1981 Misiurewicz answered it by proving that the Julia set of $E(z) = e^z$ is the entire complex plane [**M**]. Recently, Sullivan [**S**] proved the No Wandering Domains Theorem which can be used to show $J(E_\lambda) = \mathbf{C}$ for all $\lambda > 1/e$. He showed that we need only look at the behavior of certain "singular" orbits of a function to learn about its Julia set.

There are two distinct types of singular points for a complex analytic function. One type is a critical point—as in calculus, the point z_0 is a critical point if $F'(z_0) = 0$. Its image $F(z_0)$ is called a critical value. The other type of singular point is an asymptotic value. A point is called an asymptotic value if there exists a curve $\gamma(t)$ which tends to ∞ as $t \to t_0$ and which satisfies

$$\lim_{t \to t_0} F(\gamma(t)) = z_0.$$

Clearly, 0 is an asymptotic value for E_λ (take any curve tending to ∞ with real part tending to $-\infty$).

A version of Sullivan's Theorem states:

THEOREM 4.6. *Let F be an entire function with only finitely many critical and asymptotic values. Suppose the orbits of all critical points and asymptotic values of F tend to ∞. Then $J(F) = \mathbf{C}$.*

COROLLARY 4.7. *If $\lambda > 1/e$, $J(E_\lambda) = \mathbf{C}$.*

Let us briefly summarize the explosion in $J(E_\lambda)$. When $\lambda < 1/e$, E_λ admits two fixed points, q_λ and p_λ. The repelling fixed point p_λ has a "hair" attached: the line segment $[p_\lambda, \infty)$ lies in the Julia set. In this case, the singular value tends to q_λ under iteration and thus is blocked from escaping to ∞.

As λ approaches $1/e$, the two fixed points coalesce at $q_\lambda = p_\lambda = 1$, with the line $[1, \infty)$ still lying in $J(E_\lambda)$. 0 still does not reach this curve under iteration.

When $\lambda > 1/e$, two things occur. First, q_λ and p_λ disappear from the real line. Actually, one may check that the two fixed points appear as complex conjugate fixed points in the exterior of the unit disk. Hence these points have complex derivatives with modulus > 1, so nearby points tend to spiral away from them. In particular, this separation now allows 0 to hop onto the hair that formerly was attached to p_λ, and so $E_\lambda^n(0) \to \infty$. See FIGURE 9.

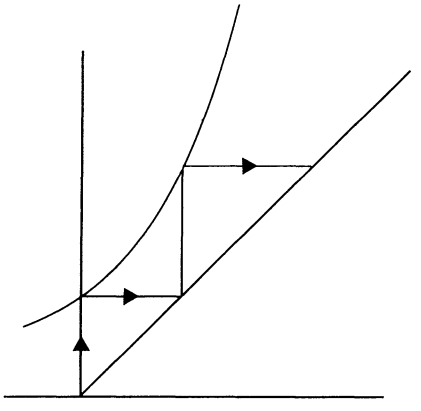


FIG. 9. $E_\lambda^n(0) \to \infty$ when $\lambda > 1/e$.

Although E_λ no longer has fixed points on the real line, there are two repelling fixed points in the complex plane. Hence, we may apply Theorem 4.5 to conclude that $J(E_\lambda)$ is the closure of the set of repelling periodic points for $\lambda > 1/e$ as well as $\lambda < 1/e$.

5. Other Chaotic Bursts. Our work in the preceding sections suggests that similar explosions occur if we vary parameters in other transcendental functions. It turns out that this is indeed the case. For example, let us consider the function $C_\lambda(z) = i\lambda \cos(z)$ where $\lambda \in \mathbf{R}$. We recall from elementary complex analysis that

$$C_\lambda(z) = \frac{i\lambda}{2}(e^{iz} + e^{-iz}).$$

If we initially restrict ourselves to the imaginary axis, we have

$$C_\lambda(iy) = \frac{i\lambda}{2}(e^{y} + e^{-y}) = i\lambda \cosh(y).$$

Hence $C_\lambda(z)$ preserves the imaginary axis just as $E_\lambda(z)$ preserved the real axis.

As in the case of E_λ, this family of maps admits a saddle-node bifurcation when $\lambda = \nu \approx .66274\ldots$. See FIGURE 10. If we consider the case when $\lambda < \nu$, we again have an attractor, q_λ, and a repeller, p_λ which now appear on the imaginary axis. Through analysis similar to that done previously we conclude that $[p_\lambda, \infty) \subset J(C_\lambda)$. We can continue as before and construct a Cantor bouquet centered about this hair. In addition, due to the fact that C_λ is even and 2π-periodic, we see that the Julia set consists of two rows of Cantor bouquet "fingers" which extend towards each other from the upper and lower half planes. The tips of these fingers can be seen in FIGURE 2a. We note that as long as $\lambda < \nu$, there will always be a strip containing the real axis which is contained in the basin of attraction of q_λ and that $J(C_\lambda)$ will be a closed, nowhere dense subset of the regions above and below this strip.

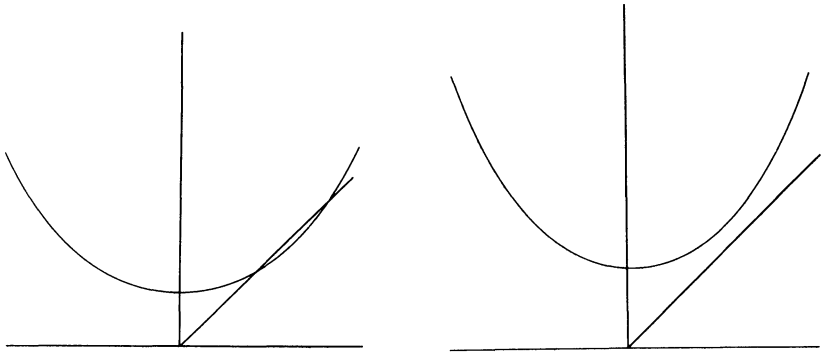


FIG. 10. The graphs of $\lambda \cosh(y)$ for $\lambda < \nu$ and $\lambda > \nu$.

Now suppose $\lambda > \nu$. We see from FIGURE 9 that all points on the imaginary axis or its 2π translates now escape to ∞ upon iteration of C_λ. Hence, each of these vertical lines is contained in $J(C_\lambda)$. Further, we notice that the real axis is mapped onto the imaginary axis making it a part of the Julia set as well. Consequently, both critical values of C_λ, $\pm i\lambda$, escape and Sullivan's Theorem guarantees that $J(C_\lambda) = \mathbf{C}$ in this case. Thus we have

THEOREM 5.1. *Let $\nu \approx .66274\ldots$ denote the λ-value for which $i\lambda \cos z$ admits a saddle-node bifurcation. If $0 < \lambda < \nu$, then $J(C_\lambda)$ is nowhere dense. If $\lambda > \nu$, $J(C_\lambda) = \mathbf{C}$.*

REFERENCES

[B] P. Blanchard, Complex Analytic Dynamics on the Riemann Sphere, B.A.M.S. vol. II, No. 1, 1984, 85–141.

[Ba] I. N. Baker, Wandering domains in the iteration of entire functions, *Proc. London Math. Soc.*, 49 (1984) 563–576.

[BR] I. N. Baker and P. Rippon, Iteration of exponential functions, *Ann. Acad. Sci. Fenn.*, Series 1A Math. vol. 9, 1984, pp. 49–77.

[Br] H. Brolin, Invariant sets under iteration of rational functions, *Ark. Math.*, 6 (1965) 103–144.

[D] R. L. Devaney, The structural instability of $\exp(z)$. *Proc. A.M.S.*, 94 (1985) 545–548.

[D1] R. L. Devaney, Julia sets and bifurcation diagrams for exponential maps, *Bull AMS*, 11 (1984) 167–171.

[DK] R. L. Devaney and M. Krych, Dynamics of $\exp(z)$, *Ergodic Theory and Dynamical Systems* 4 (1984), 35–52.

[DT] R. L. Devaney and F. Tangerman, Dynamics of entire functions near the essential singularity, *Ergodic Thy. Dynamical Syst.*, 6 (1986) 489–503.

[DH] A. Douady and J. Hubbard, Étude dynamique des polynôme complexes, *Publications Mathematiques d'Orsay*, 84–102.
[DH1] _____, Itération des polynômes quadratiques complexes, *C.R. Acad. Sci. Paris*, t.29, Serie I-1982, pp. 123–126.
[EL] A. Eremenko and M. Yu. Lyubich, Iterates of entire functions, *Dokl. Akad. Nauk SSSR*, 279 (1984), 25–27. English translation in *Soviet Math. Dokl.*, 30 (1984) 592–594.
[EL 1] _____, Structural stability in some families of entire functions, *Funk. Anal. i Prilo.*, 19 (1985) 86–87.
[F] P. Fatou, Sur l'iterâtion des fonctions transcendentes entières, *Acta Math.*, 47 (1926) 337–370.
[GK] L. R. Goldberg and L. A. Keen, Finiteness theorem for a dynamical class of entire functions, *Ergodic Theory and Dynamical Systems*, 6 (1986) 183–192.
[J] G. Julia, Iteration des applications fonctionelles, *J. Math. Pures Appl.*, (1918) 47–245.
[M] M. Misiurewicz, On iterates of e^z, *Ergodic Theory and Dynam. Syst.*, 1 (1981) 103–106.
[Ma] B. Mandelbrot, The Fractal Geometry of Nature, Freeman & Co., San Francisco, 1982.
[PR] H.-O. Peitgen and P. Richter, The Beauty of Fractals, Springer-Verlag, Heidelberg, 1986.
[S] D. Sullivan, Quasiconformal Maps and Dynamical Systems III, preprint.

LETTERS TO THE EDITOR

To the Editor:

I came out of reading the article *A New Derivation of Stirling's Approximation to n!* in the November, 1990, MONTHLY with the definite feeling that the derivation is not new. I've recently delivered classroom lectures on the derivation of the asymptotic development of $\Gamma(z)$, following Copson's *Asymptotic Expansions*, §25. The Marsaglias follow almost the same path, only using the integer n and $n!$ where Copson uses the complex variable p and $\Gamma(p)$. Their equation (3) for determining the coefficients a_n is essentially that presented by Copson. While their ultimate expression of $n!$ as a "series" is not in *Asymptotic Expansions*, it can be viewed in Copson's *Theory of Functions* at the end of essentially the same derivation.

What they have left out is any reason why their calculation should have anything more than a formal significance. They don't prove that the expansion is asymptotic, although they do mention that property, but it is the asymptoticity and the possibility of obtaining error estimates that is the goal of teaching about Stirling's formula in this century. After all,

$$e^{-1/x^2} \sim 0 + 0x + 0x^2 + 0x^3 + \cdots + 0x^n + o(x^n)$$

as $x \to 0+$, but we'll not get any idea of the value of e^{-1/x^2} by evaluating this expansion for any $x > 0$, no matter how large an n we choose. Even in a Calculus II course, where Watson's Lemma would not be at home, quotations of error bounds and illustrations of their use are appropriate and desirable.

To sum up: the authors' derivation is not new, and what they have derived is insufficient for honest numerical calculation.

Sincerely yours,

Nathaniel Grossman
Professor of Mathematics
UCLA

To the Editor:

Some readers have asked for references pertaining to my paper "A birthday present for Ramanujan," this Monthly 97 (1990) 398–400. The most accessible single reference is *Ramanujan Revisited*, Proceedings of the Centenary Conference, University of Illinois, Academic Press, 1988. The chapters by Dyson and by Garvan make wonderful reading and contain additional references.

Michael D. Hirschhorn
University of New South Wales
New South Wales, Australia

To the Editor:

In my article on the Teaching of Mathematics entitled "The l_2 matrix norm and the spectral radius" (this MONTHLY, April, 1990) there is an error which has been pointed out to me by Prof. A. K. Yeylos of Ioannina, Greece. The statement of part 1 of Theorem 1 must be changed to read as follows:

1. $\max_{\|x\|=1} (x^t Ax) = \frac{1}{2}$ *maximum eigenvalue of* $A^t + A$.

As a result, the application of this theorem in the proof of Corollary 1 must be changed also so that the second line of the proof is replaced by:

$$= \tfrac{1}{2} \textit{ maximum eigenvalue of } \left[A^t A + (A^t A)^t \right]$$

$$= \tfrac{1}{2} \textit{ maximum eigenvalue of } [2A^t A] = \rho(A^t A),$$

where the final equality results from the fact that $A^t A$ *has no negative eigenvalues.*

There is also a typographical error in the proof of Corollary 2: $\|x\|$ was printed instead of $\|x\| = 1$, as in the statement of Theorem 1.

Sincerely,

Robert A. Leslie
Agnes Scott College

To the Editor:

I wish to add a footnote to Paul Halmos' charming review article on *Progress in Mathematics*, a panorama of pure and applied mathematics in the last 75 years. In his brief report on the Bieberbach conjecture, Halmos credits Löwner with having proved the conjecture for $n = 3$. Löwner has done much more than that; he has shown that every schlicht function can be embedded in a one parameter family and has given an explicit description of all these families. Bieberbach, the editor of the Zeitschrift where Löwner's article appeared, was so enthusiastic about Löwner's idea and so convinced that it would lead to a solution in the general case that, without consulting the author, marked the paper I, the first of a series. Bierberbach was over-optimistic; it took more than 60 years before the conjecture was settled, but Löwner's representation and point of view played a decisive part in de Branges's proof.

With best regards,

Peter D. Lax
New York University

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMON

Algebraic Characterization of Compact Abelian Groups

M. V. DESHPANDE
Department of Mathematics, Indian Institute of Technology, Bombay-400076, India

SUDHIR R. GHORPADE
Department of Mathematics, Purdue University, West Lafayette, IN 47907

The well-known Banach-Stone theorem gives a nice algebraic characterization of compact Hausdorff topological spaces. One might wonder if the topological spaces are also endowed with a (compatible) group structure whether one can obtain a similar characterization for them to be the "same", i.e., to be topologically isomorphic. The purpose of this note is to give one such characterization which doesn't seem to have made its way into the literature yet. We begin by stating the result of Banach [**1**, p. 170] and Stone [**7**, p. 469] in a somewhat different form than the traditional version.

THEOREM 1. *Let X and Y be compact Hausdorff topological spaces. Then the following statements are equivalent.*

(1) *X and Y are homeomorphic.*
(2) *There exists an algebra isomorphism of $C(Y)$ onto $C(X)$.*

Here $C(X)$ (resp: $C(Y)$) denotes the space of complex valued continuous functions on X(resp: Y). These spaces clearly form an algebra (i.e., a vector space which is also a ring such that the scalar multiplication behaves nicely with the ring multiplication) over $\mathbb{C}$. In the traditional version the condition (2) requires a linear isometric isomorphism instead of an algebra isomorphism and proofs of this version could be found in [**2**, p. 442] or [**5**, p. 153]. Our proof of the result for topological groups would include a proof of Theorem 1 that is essentially as in [**4**] but does not use the language of Banach algebras and instead depends only on the following elementary fact, the proof of which is left as an exercise.

LEMMA A. *If X is any compact topological space and M is any maximal ideal of the ring $C(X)$, then there exists a unique $u \in X$ such that $M = \{f \in C(X): f(u) = 0\}$.*

We now recall the notion of a character on a topological group G which by definition is a continuous homeomorphism from G into the "circle group" $\Gamma = \{z \in \mathbb{C}: |z| = 1\}$. The characters of G form a subgroup of $C(G)$ which is usually denoted by $\hat{G}$. We will use the following basic

THEOREM 2. *Characters of a locally compact (Hausdorff) abelian group G separate the points of G, i.e.,*

$$x \in G,\ y \in G,\ x \neq y \Rightarrow \text{there exists } \chi \in \hat{G} \text{ such that } \chi(x) \neq \chi(y).$$

For a proof of this theorem we refer to ([**4**], p. 120). Let us also remark that if G is a compact group then $\hat{G}$ is discrete. Without further ado let us state and prove our main

THEOREM 3. *Let G and H be compact (Hausdorff) abelian groups. Then the following statements are equivalent.*

(1) *G and H are topologically isomorphic.*

(2) *There exists a character preserving algebra isomorphism of $C(H)$ onto $C(G)$ (i.e., $\hat{H}$ is mapped into $\hat{G}$).*

Proof. If $h\colon G \to H$ is a topological isomorphism (resp: homeomorphism) then it is easy to see that the map $T\colon C(H) \to C(G)$ defined by $T(g)(x) = g(h(x))$ for $g \in C(H)$, and $x \in G$ gives a character preserving algebra isomorphism (resp: algebra isomorphism). Conversely if $T\colon C(H) \to C(G)$ is any algebra isomorphism then we will show that T is given by $g \mapsto g \circ h$ for some homeomorphism $h\colon G \to H$ which is a topological isomorphism if T is character preserving. Given any $x \in G$ let M_x denote the ideal $\{g \in C(H)\colon T(g)(x) = 0\}$. Now M_x is clearly a maximal ideal being the kernel of the homomorphism $g \mapsto T(g)(x)$ whose image is the field $\mathbb{C}$. Hence by Lemma A there exists a unique $y = h(x) \in H$ such that $M_x = \{g \in C(H)\colon g(y) = 0\}$. In fact $T(g)(x) = g(y) = g(h(x))$ for every $g \in C(H)$ since $g - g(y)1_H$ is always in M_x. Similarly, given any $y \in H$ the ideal $\widehat{M_y} = \{f \in C(G)\colon T^{-1}(f)(y) = 0\}$, is a maximal ideal and we can find a unique $x = \hat{h}(y) \in G$ such that $\widehat{M_y} = \{f \in C(G)\colon f(x) = 0\}$. It is easy to verify that the maps h and $\hat{h}$ defined this way are inverses of each other, i.e., $h \circ \hat{h} = \mathrm{id}_H$, and $\hat{h} \circ h = \mathrm{id}_G$. Thus h is bijective. Now given any $x_0 \in G$, by Urysohn's lemma, there exists a neighbourhood V of $h(x_0)$ and some $\varphi \in C(H)$ such that $\varphi(h(x_0)) = 1$ and $\varphi(u) = 0$ for all $u \in H \setminus V$ (note that H is normal being compact and Hausdorff). Since the set $U = \{x \in G\colon T(\varphi)(x) \neq 0\}$ is an open subset of G such that $x_0 \in U$ and $h(U) \subseteq V$, it follows that h is continuous. Now both G and H are compact Hausdorff, and $h\colon G \to H$ is a continuous bijection, hence we conclude that h is a homeomorphism. This proves Theorem 1. Now assume that T preserves characters. Let $x_1 \in G, x_2 \in G$ be given. We have to show that $h(x_1 + x_2) = h(x_1) + h(x_2)$, but this is quite easy—thanks to Theorem 2. So let $\chi \in \hat{H}$ be any character. Then

$$\chi(h(x_1 + x_2)) = (\chi \circ h)(x_1 + x_2) = T(\chi)(x_1 + x_2)$$

and since $T(\chi)$ is a character, the expression on the left equals

$$T(\chi)(x_1)T(\chi)(x_2) = \chi(h(x_1))\chi(h(x_2)) = \chi(h(x_1) + h(x_2)).$$

This proves that h is a homomorphism and hence a topological isomorphism. □

REMARK. Note that an argument similar to the one used above together with Theorem 8 on page 442 of [2] would show that if G and H are as in Theorem 3 and if T is a linear isometry of $C(H)$ onto $C(G)$ which preserves characters then G and H are topologically isomorphic and in fact, there exists a topological isomorphism $h\colon G \to H$ and a character $\chi\colon G \to \Gamma$ such that $T(g) = \chi(g \circ h)$ for every $g \in C(H)$. Conversely any such T is clearly a character preserving linear isometry of $C(H)$ onto $C(G)$.

Example. Let us consider one of the simplest examples of a compact abelian group, namely the circle group $\Gamma = \{z \in \mathbb{C}\colon |z| = 1\}$. So we are taking $G = H = \Gamma$. Now it is easy to see that the characters of Γ are precisely the functions $z \mapsto z^n$ for $n \in \mathbb{Z}$ and the only topological isomorphisms among these are those functions for which $n = \pm 1$. Hence by the above theorem it follows that the only algebra automorphisms of $C(\Gamma)$ which preserve these countably many characters are the identity and the "conjugation" $f(z) \mapsto f(\bar{z})$. Also in view of the above remark we

see that the only character preserving linear isometries of $C(\Gamma)$ onto itself are $f(z) \mapsto z^n f(z)$ and $f(z) \mapsto z^n f(\bar{z})$ where n varies over the set of all integers.

REMARK. If G and H are locally compact (Hausdorff) abelian groups then we can obtain a result similar to Theorem 3 with $C(G)$ and $C(H)$ replaced by $C_0(G)$ (viz., the algebra of continuous functions on G which "vanish at infinity") and $C_0(H)$ respectively.

REFERENCES

1. S. Banach, Theorie des Opérations Linéaires, 2nd ed., Chelsea Pub. Co., New York, 1963.
2. N. Dunford and J. T. Schwartz, Linear Operators, Part 1: General Theory, Interscience, New York, 1958.
3. S. R. Ghorpade, Compact Groups and Almost Periodic Functions, M.Sc. Dissertation, Indian Institute of Technology, Bombay, 1984.
4. R. Larsen, Banach Algebras, Marcel Dekker Inc., New York, 1973.
5. J. Lindenstrauss and L. Tzafriri, Classical Banach Spaces, Springer-Verlag, Lecture Notes in Mathematics no. 338, New York, 1973.
6. L. Loomis, An Introduction to Abstract Harmonic Analysis, D. Van Nostrand, Princeton, N.J., 1953.
7. M. H. Stone, Applications of the theory of Boolean rings to general topology, *Trans. Amer. Math. Society*, 41 (1937).

First-Class Functions

G. MYERSON
Department of Mathematics, Macquarie University, New South Wales, Australia 2109

In a recent issue of this MONTHLY, J. Fabrykowski [1] posed the problem of finding a sequence of functions continuous on [0, 1] whose pointwise limit is finite on the rationals and infinite on the irrationals. To analyze a proposed solution to the problem, he required several facts from number theory; in particular, facts about continued fractions and diophantine approximation. In fact, there are simple solutions to Fabrykowski's problem which use no number theory; we give one below. Having removed all number theory from the answer, we proceed to banish it from the question, replacing the rationals with an arbitrary F_σ (that is, a countable union of closed sets). The main theorem is the following.

THEOREM 1. *There is a sequence of functions continuous on* [0, 1] *whose pointwise limit is finite on S and infinite on the complement of S if and only if S is an* F_σ.

I would like to thank my colleague, Béla Brindza, for insisting that we look for such a generalization, and for many fruitful discussions on the topic.

Let us first take care of the rationals. Gelbaum and Olmsted, in Example 4 of Chapter 7 of [2], give a sequence of continuous functions converging pointwise to the function f given by

$$f(x) = 0 \text{ if } x \text{ is irrational,}$$

$$f(p/q) = 1/q \text{ if } p \text{ and } q \text{ are relatively prime integers,} \qquad q > 0.$$

Taking reciprocals yields a sequence of continuous functions solving Fabrykowski's problem. The construction in [2] makes no use of continued fractions, diophantine approximation, or other number-theoretical devices. The following construction is similar to that of [2], and arguably even a bit simpler: Enumerate the rationals in

$(0, 1)$ as $r_1, r_2, \ldots$, and let f_n be the piecewise linear function whose graph passes through the points $(0, 0), (r_1, 1), \ldots, (r_n, n)$, and $(1, 0)$. If x is rational then x is 0 or 1 or r_k for some k, and $\lim_{n\to\infty} f_n(x)$ is 0, 0, or k, respectively—in any event, $\lim_{n\to\infty} f_n(x)$ is finite. If x is irrational, then for n sufficiently large the set $\{r_1, r_2, \ldots, r_n\}$ contains elements less than x and elements greater than x (since the rationals are dense in $[0, 1]$). Let $r_{l(n)}$ be the greatest of the elements less than x, and $r_{u(n)}$ the least of the elements greater than x; then $f(x) > \min\{l(n), u(n)\}$. Now $l(n)$ and $u(n)$ both go to infinity with n (again, because the rationals are dense in $[0, 1]$), so $\lim_{n\to\infty} f_n(x) = \infty$.

We believe that the problem of the preceding paragraph was first posed and solved over fifteen years ago, but we have not been able to locate a reference.

We now turn to the generalization. A function on an interval I is said to be of the first (Baire) class on I if it is the pointwise limit of a sequence of continuous functions on I. It is easy to see that there is a sequence of continuous functions whose pointwise limit is finite on a given subset S of $[0, 1]$ and infinite on the complement of S if and only if there is a function of the first class on $[0, 1]$ whose zero-set is the complement of S. For suppose $f_n, n = 1, 2, \ldots$, satisfies $\lim_{n\to\infty} f_n(x)$ finite for x is S, infinite otherwise; then $\lim_{n\to\infty} 1/(1 + |f_n|)$ is the desired function of the first class. Conversely, if f is of the first class, and $S = \{x: f(x) \neq 0\}$, let $f_n, n = 1, 2, \ldots$, be a sequence of continuous functions converging pointwise to f; then the sequence $1/(|f_n| + 1/n)$, $n = 1, 2, \ldots$, gets the job done. Thus Theorem 1 is equivalent to

THEOREM 2. *A necessary and sufficient condition for the existence of a function f of the first Baire class on $[0, 1]$ such that $\{x: f(x) \neq 0\} = S$ is that S be an F_σ set.*

Proof (of the necessity part of Theorem 2). Let f be of the first class on $[0, 1]$, and let $f_1, f_2, \ldots$ be a sequence of continuous functions converging pointwise to f. Let x be such that $f(x) > 0$; then $f(x) \geqslant 1/t$ for some positive integer t, whence $f_n(x) \geqslant 1/2t$ for all n sufficiently large. Conversely, if there is a positive integer t such that $f_n(x) \geqslant 1/2t$ for all n sufficiently large, then $f(x) > 0$. So,

$$\{x: f(x) > 0\} = \bigcup_{t=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{x: f_n(x) \geqslant \tfrac{1}{2}t\},$$

which is a F_σ. Similarly $\{x: f(x) < 0\}$ is an F_σ, so $\{x: f(x) \neq 0\}$ is an F_σ.

The sufficiency part of Theorem 2 follows from a theorem attributed by van Rooij and Schikhof [**3**, Theorem 11.12] to Lebesgue:

THEOREM 3. *Let I be an interval and let $f: I \to \mathbb{R}$. Then f is of the first Baire class on I if and only if $f^{-1}(U)$ is an F_σ for every open U in $\mathbb{R}$.*

A proof can be found in [**3**]; it is too long to reproduce here.

Proof (of the sufficiency part of Theorem 2). Let S be an F_σ set. We write $S = A_1 \cup A_2 \cup \ldots$ where the A_j are closed sets; we may assume $\phi = A_0 \subset A_1 \subset A_2 \subset \cdots$. Define f by

$$f(x) = \begin{cases} 1/k & \text{if } x \in A_k - A_{k-1}, k = 1, 2, \ldots, \\ 0 & \text{if } x \notin S. \end{cases} \tag{1}$$

Note that $f^{-1}(1/k) = A_k \cap \tilde{A}_{k-1}$ is the intersection of a closed set and an open set, hence both an F_σ and a G_δ (a G_δ set is a countable intersection of open sets). We are done if we can prove that f is of the first class.

Let U be open in $\mathbb{R}$. If U does not contain 0, then $f^{-1}(U)$ is a countable (possibly finite, possibly empty) union of sets of the form $f^{-1}(1/k)$, each of which is an F_σ; hence, $f^{-1}(U)$ is an F_σ.

If U does contain 0, then U contains all but finitely many of the numbers $1/k, k = 1, 2, \ldots$. Thus $f^{-1}(U)$ is the complement of a finite union of sets of the form $f^{-1}(1/k)$. Each $f^{-1}(1/k)$ is a G_δ, so the finite union is a G_δ, so again $f^{-1}(U)$ is an F_σ. Thus, by Theorem 3, f is of the first class.

We note that the necessity part of Theorem 2 follows immediately from applying the forward implication of Theorem 3 in the case $U = \mathbb{R} - \{0\}$; in fact our proof of necessity is essentially the proof given in [3] for the forward implication of Theorem 3.

Theorem 3 is quite elegant, but there are two drawbacks to using it here. One, if we rely on it, the proof of Theorem 2 is not self-contained. Two, it does not construct the sequence of functions converging to the first-class function in question. To overcome these drawbacks, we show how, given an arbitrary F_σ set S, we can construct the function f of Theorem 2 from scratch.

As before, let $S = A_1 \cup A_2 \cup \cdots$ be an F_σ, with $\phi \neq A_1 \subset A_2 \subset \cdots$. For x in $[0, 1]$ and A closed and non-empty we write $d(x, A)$ for $\min_{y \in A}|x - y|$, the distance from x to A.

Now let

$$f_{m,n}(x) = \max\left(0, \frac{1}{m(m+1)} - nd(x, A_m)\right),$$

m, n positive integers. We note that $f_{m,n}$ is continuous, bounded above by $1/m(m+1)$ and below by 0, and equal to $1/m(m+1)$ on A_m. As $n \to \infty$, $f_{m,n}$ approaches $(1/m(m+1))\chi_m$, where

$$\chi_m(x) = \begin{cases} 1 & \text{if } x \in A_m \\ 0 & \text{else} \end{cases}$$

is the characteristic function of A_m.

Let $f_n = \sum_{m=1}^{\infty} f_{m,n}$. The series converges uniformly in x, so f_n is continuous. Let $g = \lim_{n\to\infty} f_n$. We will show that g is in fact the function f defined by (1), thus giving f explicitly as a pointwise limit of continuous functions.

Let x be in $[0, 1]$. Then

$$g(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \sum_{m=1}^{\infty} f_{m,n}(x).$$

Now

$$|f_{m,n}(x)| \leqslant \frac{1}{m(m+1)}, \quad \text{and} \quad \sum_{m=1}^{\infty} \frac{1}{m(m+1)} < \infty,$$

so, by the dominated convergence theorem, we may interchange limit and sum; thus,

$$g(x) = \sum_{m=1}^{\infty} \lim_{n\to\infty} f_{m,n}(x) = \sum_{m=1}^{\infty} \frac{1}{m(m+1)}\chi_m(x).$$

Now if x is not in S then $\chi_m(x) = 0$ for all m so $g(x) = 0 = f(x)$. If x is in S then let x be in A_k (and thus in A_m for all $m \geqslant k$) but not in A_m for any $m < k$;

then

$$g(x) = \sum_{m=1}^{\infty} \frac{1}{m(m+1)} \chi_m(x) = \sum_{m=k}^{\infty} \frac{1}{m(m+1)} = \frac{1}{k} = f(x).$$

Hence, g is identical to the function f given by (1).

So we have removed a problem from the realm of number theory, and positioned it in real analysis, where it belongs. Or does it? The concepts central to Theorem 2—first Baire class and F_σ sets—are purely topological. There ought, then, to be a theorem of this type: let A and B be topological spaces with such and such properties, and let p be a point in B. Then a necessary and sufficient condition for the existence of a function $f\colon A \to B$ of the first Baire class such that $S = \{x \in A\colon f(x) \neq p\}$ is that S be an F_σ set. We leave it to the reader to explore the possibilities.

REFERENCES

1. J. Fabrykowski, On continued fractions and a certain example of a sequence of continuous functions, *Amer. Math. Monthly*, 95 (1988) 537–539.
2. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.
3. A. C. M. van Rooij and W. H. Schikhof, A Second Course on Real Functions, Cambridge University Press, 1982.

On Series of Partial Fractions

P. D. Barry and D. J. Hurley
Department of Mathematics, University College, Cork, Ireland

Examples of explicit expansions in complex Fourier series

$$f(\theta) = \sum_{n=-\infty}^{\infty} k_n e^{in\theta} \tag{1}$$

are commonly obtained in one of two ways, either 1(a) by putting $z = e^{i\theta}$ in an explicit expansion

$$F(z) = \sum_{n=-\infty}^{\infty} k_n z^n$$

in a Laurent or power series that we know; or 1(b) by using the Euler coefficient formula

$$k_n = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) e^{-in\theta}\, d\theta.$$

For examples of (1a) see e.g. Knopp [**2**, pp. 454–5]. For examples of 1(b) see, e.g., Hsu [**1**, pp. 54–56], O'Neil [**3**, p. 1057], or Pipes [**4**, pp. 55–58]. From the many examples in books of the more common Fourier series of the form

$$f(\theta) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos n\theta + b_n \sin n\theta,$$

expansions of the type (1) can be deduced on using the formulae

$$\cos n\theta = \tfrac{1}{2}(e^{in\theta} + e^{-in\theta}), \qquad \sin n\theta = \frac{1}{2i}(e^{in\theta} - e^{-in\theta}).$$

Suitable functions f for use in (1b) are polynomials, exponentials $e^{a\theta}$, finite series formed from sums and products of these, and sectionally continuous functions which are of the latter form in each interval of continuity.

Examples of explicit expansions in series of partial fractions of the form

$$g(\zeta) = \sum_{n=-\infty}^{\infty} \frac{k_n}{\zeta + n} \tag{2}$$

or

$$h(\zeta) = \sum_{n=-\infty}^{\infty} \frac{(-1)^n k_n}{\zeta + n} \tag{3}$$

can be obtained by methods involving the calculus of residues; see, e.g., Titchmarsh [**5**, pp. 110–113] or Whittaker and Watson [**6**, pp. 134–136].

As usually treated, 1(a), (2) and (3) on the one hand, and (1b) on the other seem to belong to different areas of mathematics, the former to complex analysis and the latter to real analysis. We provide a link whereby expansions in (1b) yield ones of type (2) and (3).

THEOREM. *If*

$$f(\theta) = \sum_{n=-\infty}^{\infty} k_n e^{in\theta},$$

then

$$\frac{i}{e^{2\pi i\zeta} - 1} \int_0^{2\pi} f(\theta) e^{i\zeta\theta}\, d\theta = \sum_{n=-\infty}^{\infty} \frac{k_n}{\zeta + n}, \tag{4}$$

and

$$\frac{1}{2\pi \sin \pi\zeta} \int_{-\pi}^{\pi} f(\theta) e^{i\zeta\theta}\, d\theta = \sum_{n=-\infty}^{\infty} \frac{(-1)^n k_n}{\zeta + n}, \tag{5}$$

when the relevant termwise integrations, over $[0, 2\pi]$ *and* $[-\pi, \pi]$, *respectively, are valid.*

These conclusions will become apparent on performing the termwise integrations. For examples, we note that the integrations on the left-hand sides in (4) and (5) can be performed for the class of functions f that we described above as being suitable for (1b).

As a first example we take f with period 2π and satisfying

$$f(\theta) = \pi - \theta, \quad \text{for } 0 < \theta < 2\pi. \tag{6}$$

Then we find that for $0 < \theta < 2\pi$,

$$f(\theta) = \sum_{n=-\infty}^{\infty}{}' \frac{1}{in} e^{in\theta},$$

where the prime denotes that the term corresponding to $n = 0$ is omitted, and so (4) yields

$$-\frac{i}{\zeta^2} - \frac{\pi}{i\zeta} \cot \pi\zeta = \sum_{n=-\infty}^{\infty}{}' \frac{1}{in} \frac{1}{\zeta + n}.$$

This gives, completely elementarily and by a general method, Euler's celebrated

expansion of $\pi \cot \pi\zeta$ in partial fractions (see [5, p. 113], [2, p. 207]). As for (6),

$$f(\theta) = -(\theta + \pi), \quad \text{for } -\pi < \theta < 0,$$

(5) gives the corresponding expansion for $\pi \operatorname{cosec} \pi\zeta$ (see [5, p. 113]).

As a second example, we let $0 \leqslant \lambda < \mu \leqslant \pi$,

$$f(\theta) = 1, \qquad \lambda < \theta < \mu,$$

$f(\theta)$ be 0 otherwise in $(0, 2\pi)$, and f have period 2π. Then we find, by (4) and (5), that

$$\frac{e^{i\mu\zeta} - e^{i\lambda\zeta}}{\zeta[e^{2\pi i\zeta} - 1]} = \frac{\mu - \lambda}{2\pi\zeta} + \sum_{n=-\infty}^{\infty}{}' \frac{1}{2\pi i n}[e^{-in\lambda} - e^{-in\mu}]\frac{1}{\zeta + n},$$

$$\frac{e^{i\mu\zeta} - e^{i\lambda\zeta}}{2\pi i\zeta \sin \pi\zeta} = \frac{\mu - \lambda}{2\pi\zeta} + \sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n}{2\pi i n}[e^{-in\lambda} - e^{-in\mu}]\frac{1}{\zeta + n}.$$

Clearly we can do an unlimited number of examples of this type. Both (4) and (5) can be subsumed under a more general formula if we integrate over $[\alpha, \alpha + 2\pi]$.

REFERENCES

1. H. P. Hsu, Fourier Analysis, Simon and Schuster, New York, 1970.
2. K. Knopp, Theory and Application of Infinite Series, 2nd ed., Blackie and Son, London and Glasgow, 1951.
3. P. V. O'Neil, Advanced Engineering Mathematics, 2nd ed., Wadsworth, Belmont, California, 1987.
4. L. A. Pipes, Applied Mathematics for Engineers and Physicists, McGraw-Hill, New York and London, 1946.
5. E. C. Titchmarsh, Theory of Functions, 2nd ed., Oxford University Press, 1939.
6. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., Cambridge University Press, 1927.

A Note on Dixon's Proof of Cauchy's Integral Theorem

PETER A. LOEB*
Department of Mathematics, University of Illinois, Urbana, IL 61801

In this note, we establish an elementary result in complex function theory and apply that result to simplify John Dixon's elegant proof [2] (also see [1] and [3]) of the general Cauchy integral theorem and formula. Dixon's proof uses local Cauchy theory. It is based on the fact that if f is analytic on a region G and γ is a closed rectifiable curve in G, then the integral

$$\int_\gamma \frac{f(w) - f(z)}{w - z}\, dw$$

(with the integrand replaced by $f'(w)$ when $z = w$) is an analytic function of z on all of G. A simple proof of this fact is given here, and we outline its application in Dixon's proof of Cauchy's integral formula and theorem.

*The author's work was supported in part by a grant from the U.S. National Science Foundation (DMS 87-02064).

Throughout this note, G will be an open set in the complex plane $\mathbb{C}$, and γ will be a closed rectifiable curve in G. We write $f \in H(G)$ if f is holomorphic, i.e., analytic, on G. We use the notation $D(z, r)$ for the disk $\{w \in \mathbb{C}: |w - z| < r\}$; we write $C(z, r)$ for the circle $\{w \in \mathbb{C}: |w - z| = r\}$. The trace of the curve γ in the complex plane is denoted by $\{\gamma\}$. We begin with our elementary result; it is an application of the Cauchy integral theorem for disks. The author is indebted to Professor John B. Conway for a simplification of its proof.

PROPOSITION 1. *Given $z \in \{\gamma\}$, there is a closed curve σ in G with $z \notin \{\sigma\}$ such that $\int_\gamma f = \int_\sigma f$ for all $f \in H(G)$.*

Proof. We assume that there is a point $\zeta \neq z$ with $\zeta \in \{\gamma\}$; otherwise the result is trivial. Pick $r > 0$ so that $D(z, 2r) \subset G$ and $\zeta \notin D(z, r)$. We will assume that γ is given by $\gamma(t)$ for $t \in [0, 1]$ and $\gamma(0) = \gamma(1) = \zeta$. By the uniform continuity of the mapping γ, $\exists \delta > 0$ such that if $s, t \in [0, 1]$ and $|t - s| < \delta$, then $|\gamma(t) - \gamma(s)| < r$. Let U be the inverse image under γ of the disk $D(z, r)$. Each component of U is an open interval in $(0, 1)$. If (a, b) is a component of U, then by continuity, $\gamma(a)$ and $\gamma(b)$ are points on $C(z, r)$. Moreover, if (a, b) contains a point t such that $\gamma(t) = z$, then $b - a \geqslant 2\delta$ since $|\gamma(b) - \gamma(t)| = r$ and $|\gamma(t) - \gamma(a)| = r$. Thus, there are only a finite number of components (a_i, b_i), $1 \leqslant i \leqslant n$, of U containing points t with $\gamma(t) = z$. We replace the curve γ on the closure $[a_i, b_i]$ of each such component with the arc on the circle $C(z, r)$ going from $\gamma(a_i)$ to $\gamma(b_i)$ in the positive direction. By Cauchy's integral theorem, applied to the disk $D(z, 2r)$, this replacement does not change the value of the integral for any $f \in H(G)$. The new path σ avoids z. □

The next result, Proposition 2, contains the application of Proposition 1 to Dixon's proof: Unlike published proofs of this result, the proof here does not need the joint continuity of the function φ defined below nor does it employ Fubini's theorem.

PROPOSITION 2. *Let f be analytic on G, and let φ be the mapping of $G \times G$ into $\mathbb{C}$ defined by setting $\varphi(w, z) = (f(w) - f(z))/(w - z)$ for $w \neq z$ and $\varphi(w, w) = f'(w)$. Then for each $z \in G$, $\varphi(\cdot, z)$ is analytic on G. Moreover, the function g defined by setting $g(z) = \int_\gamma \varphi(w, z)\, dw$ is analytic on G.*

Proof. Given $z \in G$, the function $\varphi(\cdot, z)$ is analytic on $G - \{z\}$ and continuous at z. From local Cauchy theory (e.g., [**3**], Theorem 10.14), it follows that $\varphi(\cdot, z) \in H(G)$. If $z \notin \{\gamma\}$, then

$$g(z) = \int_\gamma \varphi(w, z)\, dw = \int_\gamma f(w)/(w - z)\, dw - f(z) \cdot \int_\gamma 1/(w - z)\, dw.$$

It is easy to see that g is analytic on $G - \{\gamma\}$. Now fix a point $z \in \{\gamma\}$. By Proposition 1, we can replace γ with another curve σ that misses z, and therefore an open disk $D(z, \varepsilon)$ about z, without changing the values of g on $D(z, \varepsilon)$. It follows that g is analytic on $D(z, \varepsilon)$, and thus on G. □

To outline Dixon's proof of Cauchy's integral theorem and formula, we need to recall that the index $n(\gamma, a)$ of γ at any complex number $a \notin \{\gamma\}$ is given by $n(\gamma, a) = (1/2\pi i)\int_\gamma 1/(w - a)\, dw$. The index is an integer valued function that is constant on the components of the complement of $\{\gamma\}$ and 0 on the unbounded component. A cycle Γ is a finite set of closed, rectifiable curves $\{\gamma_i: 1 \leqslant i \leqslant n\}$. The trace of Γ, $\{\Gamma\}$, is $\bigcup_i \{\gamma_i\}$. For any function f continuous on $\{\Gamma\}$, the integral

$\int_\Gamma f = \sum_i \int_{\gamma_i} f$. Moreover, for $a \notin \{\Gamma\}$, $n(\Gamma, a) = \sum_i n(\gamma_i, a)$. We say that Γ is a cycle in an open set G if $\{\Gamma\} \subset G$.

THEOREM. *Let G be an open subset of $\mathbb{C}$, and let Γ be a cycle in G such that $n(\Gamma,a) = 0$ for each point $a \notin G$. Then we have Cauchy's Integral Formula*

$$\forall f \in H(G), \qquad \forall z \in G - \{\Gamma\}, \qquad f(z) \cdot n(\Gamma, z) = \frac{1}{2\pi i} \int_\Gamma \frac{f(w)}{w - z}\, dw,$$

and Cauchy's Integral Theorem

$$\forall f \in H(G), \qquad \int_\Gamma f(w)\, dw = 0.$$

Proof. For the Integral Formula, let $f \in H(G)$, and let φ and g be the mappings of Proposition 2. Let $H = \{z \in \mathbb{C} - \{\Gamma\}: n(\Gamma, z) = 0\}$. Then H is an open subset of $\mathbb{C}$ such that $G \cup H = \mathbb{C}$. Let h be the analytic function defined on H by setting

$$h(z) = \int_\Gamma \frac{f(w)}{w - z}\, dw \quad \forall z \in H.$$

At all points $z \in G \cap H$, we have

$$g(z) = h(z) - \int_\Gamma \frac{f(z)}{w - z}\, dw = h(z) - f(z) \cdot 2\pi i \cdot n(\Gamma, z) = h(z).$$

Therefore, we may extend the function g to an entire function by setting $g(z) = h(z)$ for $z \notin G$. Since $\lim_{z \to \infty} g(z) = 0$, $g \equiv 0$ on $\mathbb{C}$ by Liouville's theorem. It follows that for all points $z \in G - \{\Gamma\}$,

$$f(z) \cdot n(\Gamma, z) = \frac{1}{2\pi i} \int_\Gamma \frac{f(z)}{w - z}\, dw = \frac{1}{2\pi i} \int_\Gamma \frac{f(w)}{w - z}\, dw.$$

To establish Cauchy's Integral Theorem, we fix $a \in G - \{\Gamma\}$ and note that

$$\int_\Gamma f(w)\, dw = \int_\Gamma \frac{f(w) \cdot (w - a)}{w - a}\, dw = 2\pi i \cdot n(\Gamma, a) \cdot f(a) \cdot (a - a) = 0. \quad \square$$

REFERENCES

1. John B. Conway, Functions of One Complex Variable, second edition, Springer-Verlag, New York, 1986.
2. John D. Dixon, A brief proof of Cauchy's integral theorem, *Proc. Amer. Math. Soc.*, 29 (1971) 625–626.
3. Walter Rudin, Real and Complex Analysis, third edition, McGraw-Hill, New York, 1987.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

The Self-Paced Calculus Program at Berkeley

MURRAY H. PROTTER
Department of Mathematics, University of California, Berkeley, CA 94720

I. Introduction. About twenty years ago there were many experiments with self-paced courses at the beginning college level. The University of California was no exception; in 1971 physicist Robert Karplus and I introduced a self-paced course that combined first-year physics and first-year calculus. We wanted to remedy the justifiable complaint of many physics teachers that their students never seem to learn the mathematics needed for understanding principles of physics until *after* these principles are taught in the basic physics course. Physics instructors frequently must teach bits of mathematics to students who otherwise would be bewildered by the physics being presented.

By interweaving mathematics and physics we achieved in our interdisciplinary self-paced course a program that taught both subjects in a logical and consistent manner. The course was taught for two years and then dropped because combining physics and mathematics in one course put too heavy a burden on the students.

As I was preparing the materials for the joint mathematics-physics course, I realized that there might also be a place in the mathematics department for a self-paced alternative to the regular first-year calculus course. It was not difficult to adapt the mathematics portion of the joint mathematics-physics course into a straight calculus sequence, and I announced an experimental self-paced one-year calculus and analytic geometry course to begin in the fall term of 1972. The initial enrollment was 85. Students were enthusiastic about this form of instruction; word spread and within a few years enrollment grew to approximately 500, the maximum we can handle with the facilities we have available. Now, after 18 years the self-paced calculus course is firmly established in the Berkeley curriculum with about one-third of the students opting for the self-paced course and two-thirds taking the usual lecture-discussion section type of course.

2. How the self-paced course works. The school year at Berkeley consists of 30 weeks of instruction plus examination periods. I decided to divide the syllabus for first-year calculus with analytic geometry into 30 equal parts, called *segments*. A student proceeding at a normal pace would complete one segment each week, of course. An institution with an academic year with length different from ours could easily add or subtract segments without disturbing the continuity of the course. At Berkeley we use a standard calculus text and a specially prepared pamphlet which contains 30 *Study Guides*,* one for each of the 30 segments.

One faculty member is in charge of the course; however, he is aided by several teaching assistants and undergraduate tutors. There are no lectures or section

*Between 1972 and 1988, the text used in the self-paced course was Protter and Morrey, *College Calculus with Analytic Geometry*, second and third editions. Beginning in the fall of 1988 the text was Protter and Protter, *Calculus with Analytic Geometry*, fourth edition. The pamphlet is *Student Study Guide to Accompany the Self-Paced Program of Calculus with Analytic Geometry*, Jones and Bartlett, publishers (Boston).

meetings. Each study guide has (i) a set of objectives, (ii) a reading assignment in the text, (iii) one or two paragraphs providing the student with study hints, (iv) a set of required problems in the text, (v) additional average problems for the student who needs them, and (vi) additional challenging problems for the above average student.

The course requires the use of two fairly large classrooms for most of the day at least three days a week. One classroom is used exclusively for test-taking and the other for tutoring. Typical students start by reading the assignment described in Study Guide 1. After they believe they have mastered the objectives and solved enough of the problems to be convinced that they are prepared, they go to the testing room, where a teaching assistant gives them a test for Study Guide 1. Each student takes the test with normally no time constraints on how long he or she works. After finishing the test he* sits down with a teaching assistant to go over the examination. Typically there are three problems on a test; getting one problem wrong constitutes failure. (Passing the test means that, except for minor errors, all the problems are done correctly.) If the student passes he goes on to Study Guide 2. If he fails, he and the grader discuss both the reasons for failure and in what way the objectives of that study guide have not been met. Next, the student studies the material again, returning on a subsequent day to take an equivalent test for Study Guide 1. (There are at least four equivalent tests available for each Study Guide. Also, a key to every test is available to help the teaching assistant when he grades a test together with the student.) Thus a student can proceed to the next Study Guide only when he has mastered completely all preceding Study Guides.

The tutoring room is just that. Any student can drop in at any time the room is open and get individual help from a tutor. Tutors encourage the students to read the text carefully and to discuss with the tutor passages they find difficult. Of course, tutors help students with homework problems when such help is requested. Since homework is not graded, the students ask questions about such problems only when they need help in achieving the objectives of the Study Guide.

3. Grading and units credit. At Berkeley the regular first-year calculus course provides four units credit each semester. Thus, if a student completes 15 segments of the self-paced course each semester, he is credited with four units. However, the course actually is denoted a "variable credit course" and, through an arrangement with the registrar, the instructor assigns the number of credits a student receives (as well as a grade) at the end of each semester. There is a minimum of two units credit for which a student must enroll at the beginning of the term; thus he is required to complete seven or eight segments during the semester.** However, if he completes more than the minimum, he will receive credit for the actual number of segments he has completed.

The final examination, which every student is required to take, does pose a problem. Each term the faculty member must make up an examination with thirty problems—one corresponding to each segment. In fact, we announce that the number of the problem on the final examination identifies the segment from which the material has been selected. Then each student is required to answer only those problems which cover the segments he has completed during the term. However, since the final examination is restricted to three hours, there remains a difficulty

*Simply for convenience we shall, henceforth, use the term *he* rather than *he or she*.

**We make an exception for students who have already completed seven units of credit (26 or 27 segments). They are allowed to register for the one remaining unit of credit.

for those students who have completed more than ten segments during the term. For this contingency, we announce that if a student has completed eleven or more segments, he is required to answer only the ten problems which correspond to the most recent ten segments he has completed. The letter grade the student receives in the course is based on the scores on the segment tests (which are graded numerically) and the score on the final examination.

4. Record keeping. Good records are essential for a self-paced course. When the course began in 1972 we used a simple technique for keeping the extensive records we need. At the beginning of the term we assigned about 75 students to each teaching assistant. This assignment provided each student with a contact for answering questions about the course and for help with logistical problems. In addition to keeping a class list of his students, the TA recorded every segment test the students took and the grade received on each test. Each segment test is written on a sheet of typewriter-sized paper containing the examination questions; the student works the problems on that sheet (using additional sheets, if necessary). Everything is turned in to the TA so that no copies of the test are left in the hands of students. In a storage room we have a set of filing cabinets in which all student records are kept (including their tests). Since some students take several semesters to complete the course, we keep the records for a few years.

We began the use of computers for keeping records in the mid-seventies. Since each student entering Berkeley receives an identification number by the registrar, we use these numbers as the basis for our data bank. Currently we have a computer program and a terminal in the test-taking room so that when a student wishes to take a test he identifies himself; the TA then brings up his record on the computer screen. The record shows which segment tests the student has taken and his grades on all of them. Then the TA enters into the computer which test he is now giving to the student. After the test is completed, the grade is recorded and the examination paper is kept for the files. Periodically a clerk files all the test papers in the storage room.

5. Who should take self-paced calculus? It is easy to describe the students who should *not* take a self-paced course. We are all familiar with those students who don't begin to study until the instructor in the course announces an hour exam. Since in a self-paced course there is never any external pressure to do anything, a student who needs a deadline to get started in his work will wait until the last two or three weeks of the semester before doing anything at all. Such students invariably fail the course. Since it is difficult to prevent all students of this type from registering in the course we therefore instituted what is known as the "three-week rule." This rule restricts the number of segment tests a student may take during the last three weeks of the semester. The rule has a practical purpose as well as a psychological one, since without it the TAs and tutors would be greatly overworked near the end of the semester. While self-pacing allows the student the freedom to work in an irregular way, it does not allow him to bunch all of his test-taking toward the end of the term.

On the other hand, self-pacing is suitable for many kinds of students. In addition to those students who are turned off by large lectures and who wish to work at their own speed, there are also two other groups of students who profit from self-pacing. First, there are those students who have had some calculus in high school but who have either failed the AP test or believe that they don't know the subject well enough to skip first-year calculus entirely. Such students are bored in the large calculus lectures and believe (often mistakenly) that they already know

much of the material. If they really do know the subject, they can go through the self-paced course quickly and, as they pass all 30 tests, acquire the confidence and understanding they need in order to go on. If they don't understand the subject, they find that out quickly as they fail the tests and have to repeat them. Because many of these students are bright but are the victims of their high school courses, self-pacing is a boon to them. The second class of students for whom self-pacing is useful are those who have the ability to learn mathematics but are slow at it. In a regular course such students will get about a C and thus learn a random 70% of the material. However, in a self-paced course, the same students will learn *thoroughly* the first 70% of the course, get a good grade, but earn three units credit instead of four. As a bonus the student learns to read mathematics, gains confidence, and comes to appreciate the positive reinforcement that goes with passing the segment tests.

There are additional reasons for the great appeal of the self-paced program. For example, if an instructor in another course announces a term paper or an hour exam, the student can easily plan to do his math at times that don't interfere with his other deadlines. Perhaps the most remarkable difference between the standard lecture course and the self-paced course is the student reaction. A large majority of students dislike the lecture-discussion type of course; however, almost all students who choose the self-paced course are enthusiastic about it.

All students who take sophomore mathematics at Berkeley do so in the lecture-discussion format. From time to time we have compared the performance in sophomore mathematics of students who took the first-year lecture classes with those who took the self-paced course. Statistically, the two groups do equally well.

There is another way in which the self-paced course is helpful. Freshmen entering Berkeley can choose to take a weaker calculus course, one designed for students planning to study economics, business administration, or biology. This course, called Math 16AB, usually has an enrollment of more than 1,000 students, but it has the feature that once a student completes this course he is not qualified to go on and take any additional mathematics courses. As one would expect, each year some of these students wish to go on in mathematics. They have two choices: they can go back and take the entire regular calculus course (called Math 1AB) over again, or they can register for the self-paced course (Math 1S) and pass a modest number of segment tests selected by the mathematics department so that their background for second-year mathematics will be adequate. The students who select this last option are grateful indeed, for the self-paced course saves them considerable time and work.

Since there are many students for whom a self-paced course would be disastrous, it is important to emphasize that self-pacing cannot replace entirely the regular lecture-discussion section course. However, institutions that have several sections of calculus are in a good position to offer students the choice of two ways to learn the subject; small colleges with fewer than 50 students taking calculus should not attempt self-pacing.

6. Cost. It is difficult to assess whether or not a self-paced calculus course is more expensive than a regular lecture-discussion section course. At Berkeley we use a regular faculty member (usually a full professor) to lecture to 250 students. There is the expense of involving four faculty members to staff the first-year lecture-discussion course, which has an enrollment of 1,000 students. The self-paced course requires one faculty member for 500 students. However, all the students in the lecture-discussion section course get four units credit while the credit for the students in the self-paced course averages between three and four units. Both

courses use about the same number of teaching assistants per student. In the lecture-discussion section course we employ readers (upper-division students and some graduate students) to grade the homework that all students are required to turn in. In the self-paced course we employ tutors (upper-division students and some graduate students) on an hourly basis to help students, as described earlier.

There is also the problem of the efficient use of space. For the lecture course we use four large lecture halls three days a week for one hour and dozens of small rooms twice a week for one hour. In the self-paced course, two rooms are used all day at least three days a week. There is also a storage room. Several people here have tried to analyze this information with respect to cost and have decided that self-pacing is slightly more expensive but worth the outlay. Each institution would have to examine its own data to decide which method is more costly.

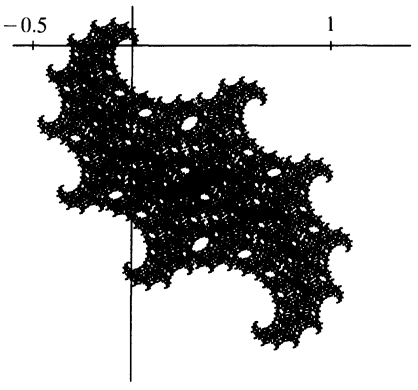
7. Other courses. Every beginning student at Berkeley who wishes to elect a mathematics course must take a diagnostic test in elementary algebra and trigonometry. Those who pass the test can choose the standard calculus course, Math 1AB (lecture-discussion section), Math 1S (self-paced), or Math 16AB, the weaker course described earlier. Those who fail the test must take a one-semester precalculus course, called Math P. After the success of the self-paced Math 1S was fully recognized, we introduced self-paced versions of Math P and Math 16AB. These courses have been operating successfully for several years, so that all our beginning mathematics students have the option of taking a lecture-discussion section course or a self-paced course. I believe that all lower division courses with large enrollments should give each student the option of attending large lectures and small discussion sections or of learning the subject at his own speed. I anticipate that the second-year mathematics course at Berkeley, which has large enrollments every term, will soon offer students these two options. It may develop that the self-paced second-year course will be even more popular than the first-year, self-paced courses are.

Number Systems With a Complex Base: a Fractal Tool for Teaching Topology

DANIEL GOFFINET
9 Rue de la Republique, St. Etienne, 42100 France

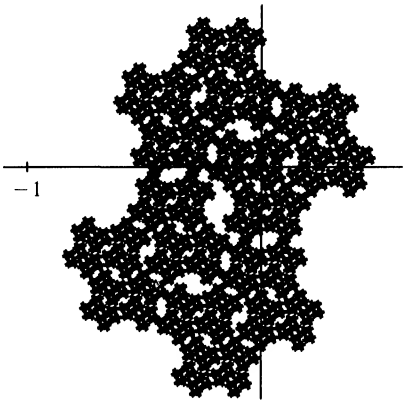
It so happens that I teach a first course in topology and a course in programming to the same students. Number systems other than the decimal one have proven to be a useful tool in the two areas. The number system with a complex base and digits 0 and 1 that is described in this note illustrates many topological notions and enthralled most of my students. An historical survey of the idea of using nonreal bases can be found in [**1**, pp. 188–190].

I gave them the program to play with, without any specific questions, in the hope of helping them to formulate and answer their own questions. Because they were only beginners in topology they mostly asked about programming details rather than the questions on connectivity that I had in mind. Still, I think it would make students with some topological background understand the necessity of the so-called "abstract" definitions.



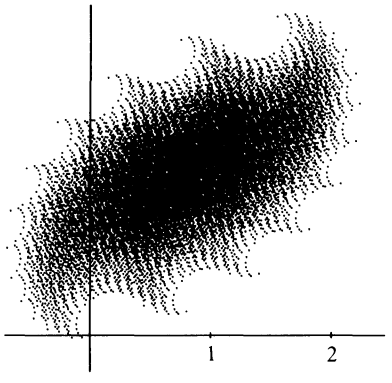
$b = 0.65 - 0.3i$

FIG. 5.



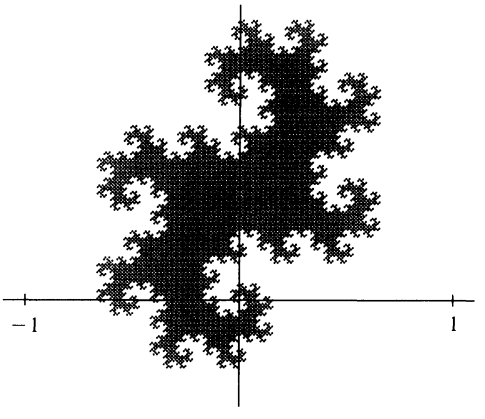
$b = -0.697e^{2\pi i/5}$

FIG. 6.



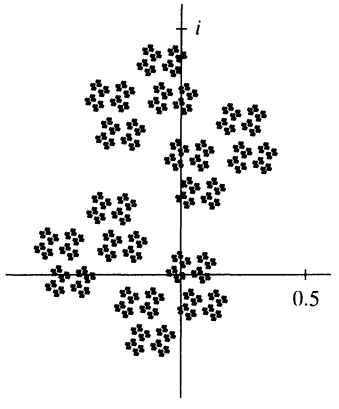
$b = 0.8 + 0.2i$

FIG. 7.



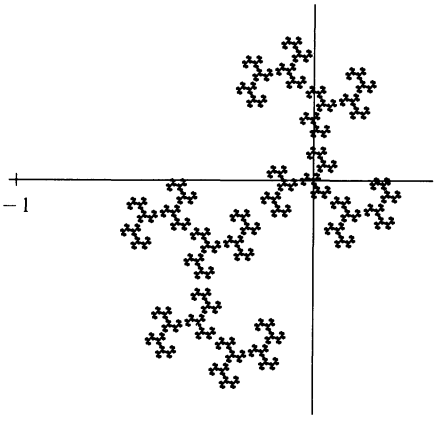
$0.5 + 0.5i$

FIG. 8.



$b = 0.2 + 0.6i$

FIG. 9.



$b = -0.62e^{2\pi i/7}$

FIG. 1–10. Some examples of $\mathfrak{B}$, where the images are produced by considering only sets A arising as subsets of $\{1, \ldots, 15\}$ (except for FIG. 9 which uses subsets of $\{1, \ldots, 13\}$).

$\mathscr{P}$ is complete. If A_n is a Cauchy sequence in $\mathscr{P}$ then for every $p \in \mathbb{N}^+$ the sets $A_n \cap [1, p]$ are eventually constant, which yields the limit.

$\mathscr{P}$ is totally bounded. For every p, $\mathscr{P}$ is a finite union of sets with diameter smaller than 2^{-p}, namely the 2^p sets obtained by considering, for each subset of $[1,p]$, all its supersets in $\mathscr{P}$. Alternatively, compactness follows from the fact that the symmetric-difference topology coincides with the product topology obtained by considering $\mathscr{P}$ as $\{0, 1\}^{\mathbb{N}^+}$ (where $\{0, 1\}$ is given the discrete topology).

If we define φ_b: $\mathscr{P} \to \mathbb{C}$ by $\varphi_b(A) = \Sigma\{b^n: n \in A\}$, then φ_b is continuous; hence, for each b, $\mathfrak{B} = \varphi_b(\mathscr{P})$ is compact. Another proof of compactness appears at the beginning of §3.

We will use the compactness of $\mathfrak{B}$ to prove that $\mathbb{C}$ is the union of copies of the set $\mathfrak{B}$ in the case that $b = (1 + i)/2$ (see FIG. 8). The easy way to see this is to look at [**1**, §4.1]. But you can do it yourself with a pencil and graph paper. Because the set is in some ways a counterpart to $[0, 1]$ (which is $\mathfrak{B}$ for $b = 1/2$), let us first look at the counterpart of $\mathbb{N}$ (which is obtained from binary expansions in $[0, 1]$ by replacing each $1/2$ by 2 and considering only finite expansions). Because b is now $(1 + i)/2$, $b = 1/B$ with $B = 1 - i$. Taking this B as base, and observing that $B^8 = 16$, leads us to look at $\mathscr{C}_7 = \{\Sigma\{B^k: k \in C\}: C \subseteq \{0, 1, 2, 3, 4, 5, 6, 7\}\}$.

We start with the baby dragon $\mathscr{C}_3 = \{\Sigma\{b^k: k \in C\}: C \subseteq \{0, 1, 2, 3\}\}$. When 4 is added to the set of admissible exponents, $\mathscr{C}_3$ becomes $\mathscr{C}_4$, which consists of two copies of $\mathscr{C}_3$. Continue until reaching $\mathscr{C}_7$, which consists of 16 $\mathscr{C}_3$s (see Figs. 11–12). Now it is easy to check that another copy of $\mathscr{C}_7$ translated 16 units rightward will fit with the first copy as in a jigsaw puzzle. Repeating yields a strip of $\mathscr{C}_7$s (see Fig. 13). Again, inspection shows that translating this doubly infinite strip by $16i$ yields a strip that fits with the first one. Thus every Gaussian integer (i.e., element of $\mathbb{G} = \mathbb{Z} + \mathbb{Z}i$) may be written uniquely in the form $c + 16(p + qi)$ with $c \in \mathscr{C}_7$ and $p, q \in \mathbb{Z}$. In short: $\mathbb{G} = 16\mathbb{G} + \mathscr{C}_7$.

Now, let us return to proving that $\mathbb{C}$ can be covered by $\mathbb{G}$-translations of $\mathfrak{B} = \{\Sigma_{k \geq 1} a_k b^k: a_k \in \{0, 1\}\}$. Let $\mathfrak{B}_n = \{\Sigma_{1 \leq k \leq 8n} a_k b^k: a_k \in \{0, 1\}\}$, with $\mathfrak{B}_0 = \{0\}$. Then $\mathfrak{B}_1 = b^8 \mathscr{C}_7 = \mathscr{C}_7/16$. Because $\mathfrak{B}_1/16 = \{\Sigma_{9 \leq k \leq 16} a_k b^k: a_k \in \{0, 1\}\}$, we have that $\mathfrak{B}_2 = \mathfrak{B}_1 + (1/16)\mathfrak{B}_1 = \mathfrak{B}_1 + (1/16^2)\mathscr{C}_7$. Similarly $\mathfrak{B}_n = \mathfrak{B}_{n-1} + (1/16^n)\mathscr{C}_7$. Assume inductively that $\mathfrak{B}_{n-1} + \mathbb{G} = \mathbb{G}/16^{n-1}$ (clearly, $\mathfrak{B}_0 + \mathbb{G} = \mathbb{G}$). Then $\mathfrak{B}_n + \mathbb{G} = (\mathfrak{B}_{n-1} + \mathbb{G}) + (1/16^n)\mathscr{C}_7 = (1/16^{n-1})\mathbb{G} + (1/16^n)\mathscr{C}_7 = (1/16^n)(16\mathbb{G} + \mathscr{C}_7) = \mathbb{G}/16^n$. Since $\mathfrak{B} + \mathbb{G} \supseteq \mathfrak{B}_n + \mathbb{G}$, $\mathfrak{B} + \mathbb{G}$ contains all points in a grid of mesh $(1/16)^n$. It follows that for any square there is a finite union of $\mathbb{G}$-translates of $\mathfrak{B}$ (a compact set) which is dense in the square. Hence, $\mathfrak{B} + \mathbb{G}$ contains the square, and so $\mathfrak{B} + \mathbb{G} = \mathbb{C}$.

As with the usual base-10 representations, we lose uniqueness when we pass to the limit; for example: $b = 1 + (b^2 + b^3 + \cdots)$.

3. Going deeper into the topology of $\mathfrak{B}$. I owe the following fixed-point idea to [**5**]. If $\mathscr{K}$ is the set of nonempty, compact subsets of the plane, the Hausdorff distance is defined as $\delta(X, Y) = \max\{\sup d(x, Y), \sup d(y, X)\}$; then $(\mathscr{K}, \delta)$ is a complete space [**2**, p. 56]. I now define T_0 by $T_0(z) = bz$ and T_1 by $T_1(z) = b + bz$. The functions T_0 and T_1 are contractions of $\mathbb{C}$, and also clearly induce contractions of $(\mathscr{K}, \delta)$. If $T = T_0 \cup T_1$ is defined by $T(X) = T_0(X) \cup T_1(X)$, then T is also a contraction of $(\mathscr{K}, \delta)$, because we can easily prove that

$$\delta(X_1 \cup X_2, Y_1 \cup Y_2) \leq \max\{\delta(X_1, Y_1), \delta(X_2, Y_2)\}.$$

Therefore T has a unique fixed point in $\mathscr{K}$, and it is the limit of the sequence of the $T^n(K_0)$ for any K_0. Taking $K_0 = \{0\}$ we have that $T^n(K_0)$ equals those points

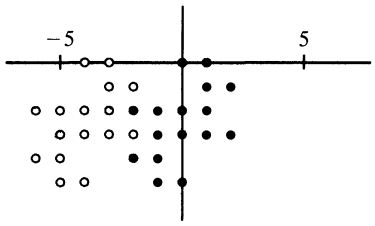


FIG. 11. Two copies of the baby dragon, $\mathscr{C}_3$, yields $\mathscr{C}_4$.

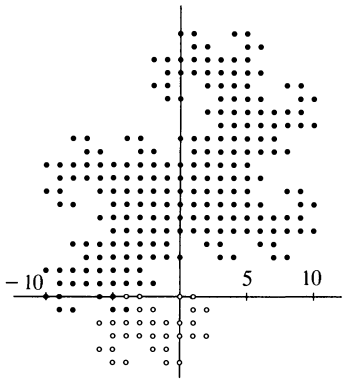


FIG. 12. The set $\mathscr{C}_7$, which is built up from 8 copies of $\mathscr{C}_4$.



FIG. 13. The Gaussian integers can be tiled with copies of the set $\mathscr{C}_7$.

in $\mathfrak{B}$ with at most n "digits" in their representation. So $\mathfrak{B}$, the limit of the sequence, is in $\mathscr{K}$ (i.e., is compact) and is a fixed point of T.

4. Connectivity. I now claim that $\mathfrak{B}$ is either connected or totally disconnected. Set $\mathfrak{B}^0 = T_0(\mathfrak{B})$ and $\mathfrak{B}^1 = T_1(\mathfrak{B})$; then $\mathfrak{B} = T(\mathfrak{B}) = \mathfrak{B}^0 \cup \mathfrak{B}^1$.

Case 1. $\mathfrak{B}^0 \cap \mathfrak{B}^1 = \varnothing$. If $m \in \mathfrak{B}^i$, then $\mathscr{M}$, its connected component in $\mathfrak{B}$, is contained in $\mathfrak{B}^i$ ($\mathfrak{B}^0$ and $\mathfrak{B}^1$ are compact as images of $\mathfrak{B}$ by the continuous functions T_0 and T_1, respectively). But $\mathfrak{B}^i$ itself is a disjoint union of compact sets so $\mathscr{M}$ is included in some $T_jT_i(\mathfrak{B})$. By induction we can prove that $\mathscr{M}$ is a subset of the intersection of a decreasing sequence of compact sets whose diameters tend to zero. It follows that $\mathscr{M}$ is simply $\{m\}$.

Case 2. $\mathfrak{B}^0 \cap \mathfrak{B}^1 \neq \varnothing$. In this case $\mathfrak{B}$ is connected. Choose $a \in \mathfrak{B}^0 \cap \mathfrak{B}^1$, say, $a = ba_0 = b + ba_1$. Because $\mathfrak{B}$ is compact, to prove that it is connected it is enough to check that for every $z \in \mathfrak{B}$ and $\varepsilon > 0$, there is an ε-chain in $\mathfrak{B}$ from z to a. If ε is greater than the diameter of $\mathfrak{B}$ then (z, a) is such a chain. Let us show that from α-chains we can build $\lambda\alpha$-chains where $\lambda < 1$ is the modulus of b. Since $z \in \mathfrak{B}$ we have two cases:

If $z \in \mathfrak{B}^0$ then $z = bd$, and we can "α-link" d to a, then a to a_0. Now take the images by T_0 of these chains (i.e., simply multiply by b) to $\lambda\alpha$-link bd to ba to $ba_0 = a$.

If $z \in \mathfrak{B}^1$ then $z = b + bd$, and we can α-link d to a to a_1. Then take the images of these chains by T_1.

"digits" are needed to make $\mathfrak{B}$ look whole. Another example: $\mathfrak{B}_9$ is disconnected, although one cannot really tell from the image—a proof is needed. The four upper right blobs can be separated from the four bottom left ones by a strip that can be physically measured and seen to be greater than $1/50$. The modulus of b is smaller than 0.64, so with the powers greater than 13 a gap wider than $0.64^{14}/(1 - 0.64)$ (which is well under $1/50$) cannot be bridged. Therefore $\mathfrak{B}_9$ is indeed disconnected.

The set of those b for which $\mathfrak{B}$ is connected is described in [**3**], where, among other things, it is proven that $\mathfrak{B}$ is connected if $|b| > \sqrt{2}/2$. Among the $\mathfrak{B}$s shown in FIGURES 1–10, only $\mathfrak{B}_9$ and $\mathfrak{B}_{10}$ are disconnected.

8. Miscellaneous. A lot of things are left to be explored. When $\mathfrak{B}$ is connected and simply connected (see FIGS. 7 and 8), is the boundary of fractal dimension strictly greater than 1 (as it seems)? Is that boundary the image of a continuous map from $[0, 1]$ in $\mathbb{R}^2$ (as it seems)? Why does $\mathfrak{B}_5$ have $\mathfrak{B}_5$-like eggs inside it (dark gray)?

I tried to replace $\mathbb{C}$ by the algebra with $(a, b) \cdot (c, d) = (ac + bd, ad + bc)$, but got nothing worth showing. Why not?

Acknowledgement. I am grateful to the referees for many suggestions.

REFERENCES

1. D. E. Knuth, The Art of Computer Programming, volume 2, Seminumerical Algorithms, Addison-Wesley, Reading, Mass., 1981.
2. J. Dieudonné, Fondements de l'Analyse Moderne, Vol. 1, Gauthier-Villars, Paris, 1965.
3. M. Barnsley, Fractals Everywhere, San Diego, Academic Press, 1988
4. W. J. Gilbert, Fractal geometry derived from complex bases, *Math. Intelligencer*, 4 (1982) 78–86
5. J. E. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.*, 30 (1981) 713–747.

A Characterization of the Cantor Function

DONALD R. CHALICE
Department of Mathematics, Western Washington University, Bellingham, WA 98225

This note presents a characterization of the Cantor function that might be used as an alternative or addition to the development usually presented in real analysis courses. It may also be utilized to give a short program in *Mathematica* that easily generates the Cantor function and other similar functions which we call "devil's staircases" (see [**4**]).

THEOREM. *Any real-valued function $F(x)$ on $[0, 1]$ that is monotone increasing and satisfies (a) $F(0) = 0$, (b) $F(x/3) = F(x)/2$, and (c) $F(1 - x) = 1 - F(x)$, is the Cantor function.*

Before presenting the proof, recall (see [**1**]) that if we consider the closed interval $[0, 1]$ and remove the open middle third, $(1/3, 2/3)$, and next remove the open middle thirds $(1/9, 2/9)$ and $(7/9, 8/9)$ of the two remaining intervals, and then remove the open middle thirds of the remaining four intervals and so on, indefinitely, what remains is the *Cantor set C*. Alternatively, x is in C iff x has a base-3 expansion consisting only of the digits 0 and 2.

"digits" are needed to make $\mathfrak{B}$ look whole. Another example: $\mathfrak{B}_9$ is disconnected, although one cannot really tell from the image—a proof is needed. The four upper right blobs can be separated from the four bottom left ones by a strip that can be physically measured and seen to be greater than $1/50$. The modulus of b is smaller than 0.64, so with the powers greater than 13 a gap wider than $0.64^{14}/(1 - 0.64)$ (which is well under $1/50$) cannot be bridged. Therefore $\mathfrak{B}_9$ is indeed disconnected.

The set of those b for which $\mathfrak{B}$ is connected is described in [**3**], where, among other things, it is proven that $\mathfrak{B}$ is connected if $|b| > \sqrt{2}/2$. Among the $\mathfrak{B}$s shown in FIGURES 1–10, only $\mathfrak{B}_9$ and $\mathfrak{B}_{10}$ are disconnected.

8. Miscellaneous. A lot of things are left to be explored. When $\mathfrak{B}$ is connected and simply connected (see FIGS. 7 and 8), is the boundary of fractal dimension strictly greater than 1 (as it seems)? Is that boundary the image of a continuous map from $[0, 1]$ in $\mathbb{R}^2$ (as it seems)? Why does $\mathfrak{B}_5$ have $\mathfrak{B}_5$-like eggs inside it (dark gray)?

I tried to replace $\mathbb{C}$ by the algebra with $(a, b) \cdot (c, d) = (ac + bd, ad + bc)$, but got nothing worth showing. Why not?

Acknowledgement. I am grateful to the referees for many suggestions.

REFERENCES

1. D. E. Knuth, The Art of Computer Programming, volume 2, Seminumerical Algorithms, Addison-Wesley, Reading, Mass., 1981.
2. J. Dieudonné, Fondements de l'Analyse Moderne, Vol. 1, Gauthier-Villars, Paris, 1965.
3. M. Barnsley, Fractals Everywhere, San Diego, Academic Press, 1988
4. W. J. Gilbert, Fractal geometry derived from complex bases, *Math. Intelligencer*, 4 (1982) 78–86
5. J. E. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.*, 30 (1981) 713–747.

A Characterization of the Cantor Function

DONALD R. CHALICE
Department of Mathematics, Western Washington University, Bellingham, WA 98225

This note presents a characterization of the Cantor function that might be used as an alternative or addition to the development usually presented in real analysis courses. It may also be utilized to give a short program in *Mathematica* that easily generates the Cantor function and other similar functions which we call "devil's staircases" (see [**4**]).

THEOREM. *Any real-valued function $F(x)$ on $[0, 1]$ that is monotone increasing and satisfies (a) $F(0) = 0$, (b) $F(x/3) = F(x)/2$, and (c) $F(1 - x) = 1 - F(x)$, is the Cantor function.*

Before presenting the proof, recall (see [**1**]) that if we consider the closed interval $[0, 1]$ and remove the open middle third, $(1/3, 2/3)$, and next remove the open middle thirds $(1/9, 2/9)$ and $(7/9, 8/9)$ of the two remaining intervals, and then remove the open middle thirds of the remaining four intervals and so on, indefinitely, what remains is the *Cantor set C*. Alternatively, x is in C iff x has a base-3 expansion consisting only of the digits 0 and 2.

The *Cantor function* $G(x)$ may be defined as follows. First define it on C: if $x = \Sigma_i 2 \cdot 3^{-n_i}$, then $G(x) = \Sigma_i 2^{-n_i}$. The function G is monotone and has the same values at the endpoints of each removed interval, so G extends to a continuous function on $[0, 1]$ (see FIGURE 3(a)).

On $[1/3, 2/3]$, G has value $1/2$, on $[1/9, 2/9]$ and $[7/9, 8/9]$, G has values $1/4$ and $3/4$, respectively. On the intervals $[1/27, 2/27]$, $[7/27, 8/27]$, $[19/27, 20/27]$, and $[25/27, 26/27]$, the values of G are $1/8$, $3/8$, $5/8$, and $7/8$, respectively. Since G is (locally) constant on some neighborhood of every point in $[0, 1] \setminus C$, $G'(x) = 0$ almost everywhere on $[0, 1]$. (See [**1**] or [**3**].)

Proof of the theorem. First observe that in constructing the Cantor set, the removed intervals (in base 3) are as given in TABLE 1. The endpoints of any removed interval at the nth stage are found by either multiplying those from the previous stage by .1 (base 3), or by multiplying by .1 and adding .2.

TABLE 1

Step	Values of G	Removed Intervals (closures)
1	1/2	[.1, .2]
2	1/4, 3/4	[.01, .02], [.21, .22]
3	1/8, 3/8	[.001, .002], [.021, .022],
	5/8, 7/8	[.201, .202], [.221, .222]
4	1/16, 3/16	[.0001, .0002], [.0021, .0022],
	5/16, 7/16	[.0201, .0202], [.0221, .0222]
	9/16, 11/16	[.2001, .2002], [.2021, .2022],
	13/16, 15/16	[.2201, .2202], [.2221, .2222], etc.

Alternatively, any permutation of $n - 1$ 0s and 2s after the ternary point followed by a 1 or a 2 gives a removed interval at the nth stage.

Recursion is now used to characterize F: By (a) and (c), $F(1) = 1$, so $F(.1) = 1/2$. By (b) and (c), $F(.2) = 1 - F(.1) = 1/2$; (b) implies $F(.01) = F(.02) = 1/4$ and (c) implies $F(.22) = F(.21) = 3/4$; (b) then yields the correct values of $1/8$, $3/8$ on the firsts two intervals at stage 3, while (c) yields the value $F(.201) = 1 - F(.022) = 5/8$. It follows by induction that F and G agree on a dense subset of $[0, 1]$. Since F is monotone increasing and has no jump discontinuities, F is continuous (because any discontinuity of a monotone function is a jump; see [**1**, p. 129]). Thus since F is continuous and agrees with the Cantor function on a dense set, F is the Cantor function, and the proof is complete.

As an alternative proof of continuity, F may be exhibited as a uniform limit of a sequence of monotone increasing step functions $\{s_n\}$ on $[0, 1]$ with the jumps converging to 0. For example, define

$$s_1(x) = \begin{cases} 0 & \text{if } x < .1 \\ 1/2 & \text{if } .1 \leqslant x \leqslant .2 \\ 1 & \text{if } .2 < x \end{cases}$$

and inductively, define

$$s_n(x) = \begin{cases} s_{n-1}(3x)/2 & \text{if } x < .1 \\ 1/2 & \text{if } .1 \leqslant x \leqslant .2 \\ 1 - s_n(1 - x) & \text{if } .2 < x \end{cases}$$

(assuming s_n as defined on the first line). The continuity of F is then a consequence of the following lemma, which may be of independent interest.

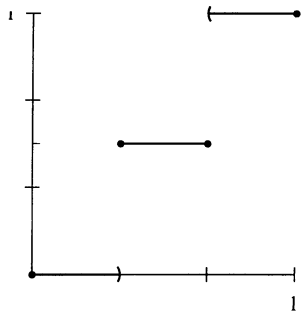


Figure 1(a) Graph of s_1.

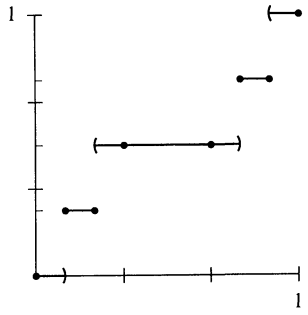


Figure 1(b) Graph of s_2.

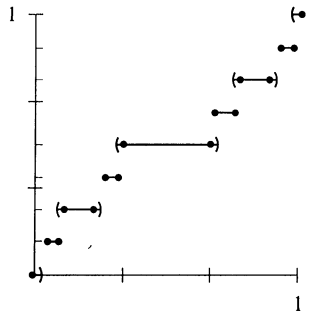


Figure 1(c) Graph of s_3.

LEMMA. *Let* $\{s_n\}$ *be a Cauchy sequence in uniform norm of step functions on* $[0, 1]$, *each with a finite number of jumps such that the heights of the jumps for* s_n *converge uniformly to* 0. *Then* s_n *converges uniformly to a continuous function* F *on* $[0, 1]$.

Proof. Pick x in $[0, 1]$ and $\varepsilon > 0$. There exists an integer N such that $n, m \geqslant N$ implies $|s_n(x) - s_m(x)| < \varepsilon/4$ for all x in $[0, 1]$, and so that the jumps of s_n are all less than $\varepsilon/4$. Now any s_n is uniformly continuous on each interval on which there is not a jump, so we can pick δ so that $|x - y| < \delta$ implies $|s_N(x) - s_N(y)| < 2\varepsilon/4$. Hence, $|x - y| < \delta$ implies $|F(x) - F(y)| \leqslant |F(x) - s_N(x)| + |s_N(x) - s_N(y)| + |s_N(y) - F(y)| < \varepsilon$. Thus F is continuous. QED

Since $n < m$ implies $|s_n(x) \neg s_m(x)| \leqslant |s_{n-1}(y) - s_{m-1}(y)|/2$ where y is one of $3x$ or $1 - 3x$, it then follows inductively that $|s_n(x) - s_m(x)| \leqslant 1/2^n$ when $n \leqslant m$. Thus $\{s_n\}$ is uniformly Cauchy and the continuity of F follows from the lemma.

Note also that F may be generated by a geometric algorithm, where we begin by defining $F(x) = 1/2$ on the interval $[.1, .2]$ and $F(0) = 0$, $F(1) = 1$, and where at each stage of the algorithm we shrink the x axis by a factor of $1/3$ and the y axis by a factor of $1/2$ and then flip the resultant graph across the line $x = 1/2$, and then again across the line $y = 1/2$. We see F as coming more into focus at each stage.

The preceding characterization was used (together with the recursive feature of *Mathematica*) to generate FIGURES 3(a)–(b) (programmed by Gerald Harnett). Other "devil's staircases" were obtained by taking any $p > 0$ and changing condition (c) to (c'): $F(1 - x) = 1 - pF(x)$, for $x \leqslant 1/3$ and condition (b) to (b'): $F(x/3) = F(x)/(p + 1)$. These generalizations arise as the cumulative distribution functions of some of the probability measures invariant under the "inverted

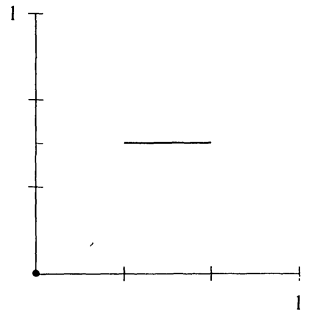


Figure 2(a).

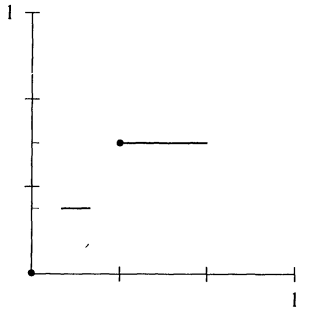


Figure 2(b)
Shrink 2(a) and superimpose on 2(a).

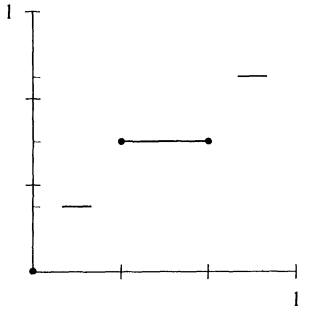


Figure 2(c)
Double flip 2(b) and superimpose on 2(b).

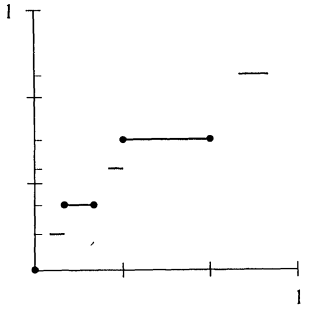
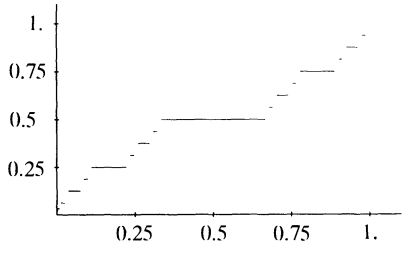
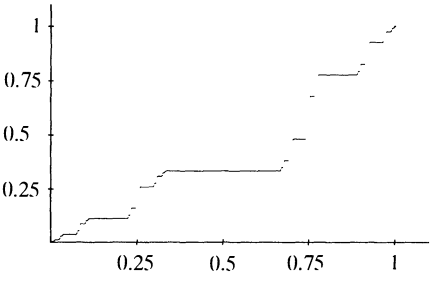


Figure 2(d)
Shrink 2(c) and superimpose on 2(c).



The Cantor function (p = 1).
FIG. 3a



The Cantor function (p = 2).
FIG. 3b

V" transformation discussed by Mandelbrot (see [2, 4]). Finally, by letting p take on, for example, the values .01, .1, .3, .5, 1, 4, 10, and 100, one can create a movie in *Mathematica* that shows the Cantor function "stressed" by the varying of the parameter p (see FIGURE 3).

REFERENCES

1. R. P. Boas, Jr., A Primer of Real Functions, Carus Mathematical Monographs, No. 13, Wiley, Rahway, New Jersey, 1960.
2. D. R. Chalice, Devil's staircases invariant under the inverted V function, to appear.
3. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.
4. B. B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman, New York, 1983.

Parametric Differentiation Revisited

L. R. Bragg
Department of Mathematical Sciences, Oakland University, Rochester, MI 48309

A number of now-ancient texts in advanced calculus employed the method of parametric differentiation to deduce evaluations for a variety of complicated definite integrals (e.g., [**4, 7**]). In his autobiography [**3**], R. P. Feynman mentioned how he frequently used this approach when confronted with difficult integrations associated with mathematical and physical problems. He referred to that method as "a different box of tools." Most modern texts either ignore the subject or provide only one or two examples to illustrate theorems on the uniform convergence of improper integrals. (See [**8**] for a more recent treatment of integrals.) However, the method of parametric differentiation has a much broader scope than this. It can, in fact, be introduced into the elementary calculus to simplify the treatment of partial fractions with repeated factors, to carry out integrations ordinarily handled by parts, and to treat integrals that use substitutions. When tied in with other mathematical machinery (e.g., complex variables), it can be used to deduce formulas for special functions and to solve problems in differential equations that involve iterated operators. Finally, the method can be conveniently used in conjunction with mathematical software packages that handle symbolic differentiations.

In this note, we give three examples from calculus to illustrate the flexibility of this approach at the elementary level. One of the integrals treated is connected with the t-distribution. We also note, without providing a complete discussion, how this technique can be used to evaluate the Riemann zeta function at $2p$, $p = 1, 2, \ldots$, and how it can be used in differential equations. A more extensive set of applications is available from the author. In the work to follow, we call upon the Leibniz product rule for differentiation as well as theorems that permit interchanging orders of differentiation and integration (see [**5, 6**]). We also need to make frequent use of the formula

$$D_a^{n-1}(x+a)^{-c} = (-1)^{n-1}c(c+1)\cdots(c+n-2)(x+a)^{-(c+n-1)}. \quad (1)$$

When $c = 1$, the right-hand side reduces to $(-1)^{n-1}(n-1)!(x+a)^{-n}$. We omit constants of integration from the indefinite integrations.

A Partial Fraction Decomposition. Suppose we wish to obtain a partial fraction decomposition of the rational function $(x+2)^{-m}(x+1)^{-n}$ where m and n are positive integers greater than 1. We associate with this the parameterized function $h(x) = (x+a)^{-1}(x+b)^{-1}$, $a > 0$, $b > 0$, $a \neq b$. By the usual algebraic decomposition method, we find

$$h(x) = (b-a)^{-1}(x+a)^{-1} + (a-b)^{-1}(x+b)^{-1}. \quad (2)$$

Upon differentiating (2) $m-1$ times with respect to a and $n-1$ times with respect to b and applying the Leibniz rule and (1) to each of the summands on

(2)'s right-hand side, we can show, after simplification, that

$$(x+a)^{-m}(x+b)^{-n} = \frac{1}{(n-1)!}\sum_{j=0}^{m-1}(-1)^j\frac{(n+j-1)!}{j!}(b-a)^{-n-j}(x+a)^{-(m-j)} + \frac{1}{(m-1)!}\sum_{j=0}^{n-1}(-1)^j\frac{(m+j-1)!}{j!}(a-b)^{-m-j}(x+b)^{-(n-j)}. \quad (3)$$

The required decomposition follows by replacing a in (3) by 2 and b by 1.

A brief test was carried out on Maple 4.2 for the Macintosh to compare the speed of Maple's built in partial fraction decomposition procedure with the speed of the parametric method. The preceding example was selected with m and n approximately 20. The differentiation method ran somewhat faster with less garbage collecting. Programs are now being developed for implementing this parametric decomposition approach on the HP-28S calculator.

An Indefinite Integral. The indefinite integral

$$I(x,a,p+n) = \int (a+x^2)^{-p-n}\,dx, \quad (4)$$

$a > 0$, $p > 0$, n a positive integer, appears frequently in applications, in particular in probability and statistics. Calling upon (1), it follows that

$$I(x,a,p+n) = \frac{(-1)^n}{p(p+1)\cdots(p+n-1)}D_a^n I(x,a,p). \quad (5)$$

If the integral $I(x,a,p)$ can be evaluated, then (5) provides a convenient way for computing $I(x,a,p+n)$. For example, if $p = 3/2$, then

$$I\left(x,a,\frac{3}{2}\right) = x(a+x^2)^{-1/2}\cdot a^{-1} \quad (6)$$

and the right-hand side is well suited for applying the Leibniz rule and (1) to evaluate (4).

This can be used to quickly obtain a finite sum formula for the t-distribution with $2m$ degrees of freedom. That distribution is defined by the improper definite integral

$$T(2m,y) = C_{2m}\int_{-\infty}^{y}\left(1+\frac{\sigma^2}{2m}\right)^{-m-1/2}d\sigma,$$

where $C_{2m} = \Gamma(m+\frac{1}{2})/(\sqrt{2m\pi}\,\Gamma(m))$. Analogous to (6), we have

$$\int_{-\infty}^{y}\left(a+\frac{\sigma^2}{2m}\right)^{-3/2}d\sigma = ya^{-1}\left(a+\frac{y^2}{2m}\right)^{-1/2} + \sqrt{2m}\,a^{-1/2}.$$

If we differentiate both sides of this $m-1$ times with respect to a, evaluate at $a = 1$, and take into account the constant C_{2m}, we can establish (see [1])

$$T(2m,y) = \frac{1}{2\sqrt{2m}}\left(\frac{y}{\sqrt{\pi}}\sum_{j=0}^{m-1}\frac{\Gamma(j+\frac{1}{2})}{j!}\left(1+\frac{y^2}{2m}\right)^{-(j+\frac{1}{2})} + \sqrt{2m}\right). \quad (7)$$

The case of the t-distribution with $2m+1$ degrees of freedom can be treated similarly.

An Integration by Parts. We wish to evaluate $\int x^n e^x \cos x\,dx$. This quadrature is usually handled by a repeated integration by parts. Let us, however, associate with this the complex antiderivative $\int e^{cx}\,dx = e^{cx}c^{-1}$ where $c = a + ib$. Then

$$\int x^n e^{cx}\,dx = D_c^n \int e^{cx}\,dx = D_c^n(e^{cx}c^{-1})$$

$$= \sum_{j=0}^{n} (-1)^j \frac{n!}{(n-j)!} x^{n-1} e^{cx} c^{-(j+1)}. \tag{8}$$

Writing c in the polar form $c = \rho e^{-i\theta}$ with $\rho = (a^2 + b^2)^{1/2}$ and $\theta = \tan^{-1}(b/a)$, (8) becomes

$$\int x^n e^{cx}\,dx = \sum_{j=0}^{n} (-1)^j \frac{n!}{(n-j)!} \frac{x^{n-j} e^{[a+i(b-(j+1)\theta)]x}}{(a^2+b^2)^{(j+1)/2}}. \tag{9}$$

Finally, if we replace a and b by 1 in (9) and equate the real parts of its two members, we find

$$\int x^n e^x \cos x\,dx = e^x \sum_{j=0}^{n} (-1)^j \frac{n! x^{n-j}}{(n-j)!\,2^{(j+1)/2}} \cos\left[x - (j+1)\frac{\pi}{4}\right]. \tag{10}$$

Other Uses.

(i) *Evaluation of* $\zeta(2p)$. The following equality for meromorphic functions is developed in complex variables:

$$\sum_{n=1}^{\infty} (n^2 + z^2)^{-1} = \frac{\pi \cosh \pi z}{2z \sinh \pi z} - \frac{1}{2z^2}. \tag{11}$$

From the fact that $\cosh \pi z/\sinh \pi z = 1 + 2/(e^{2\pi z} - 1)$, the right-hand side of (11) can be replaced by $\pi/2z - 1/(2z^2) + (1/2z)\sum_{j=0}^{\infty} B_j (2\pi z)^j/j!$ in which B_j denotes the jth Bernoulli number [**1**] with $B_{2m+1} = 0$, $m \geqslant 1$. If we replace z^2 in (11) by a positive parameter a and then differentiate (11) p times with respect to a at $a = 0$, we obtain the familiar formula

$$\zeta(2p) = \sum_{n=1}^{\infty} \frac{1}{n^{2p}} = \frac{(-1)^{p-1}(2\pi)^{2p}}{2(2p)!} B_{2p}, \qquad p = 1, 2, \ldots. \tag{12}$$

(ii) *Differential equations with repeated factors.* Let $L(x, D_x)$ denote a linear ordinary differential operator in x and D_x and suppose that $y(x, a)$ is a solution of the nonhomogeneous equation $(L(x, D_x) - a)y(x, a) = f(x)$. Then, with a little work, the reader can show that the function $y_n(x) = (D_a^{n-1} y(x, a)|_{a=0})/(n-1)!$ satisfies the iterated nonhomogeneous differential equation $[L(x, D_x)]^n y_n(x) = f(x)$. For example, if $L(x, D_x) = D_x - \lambda$, λ a constant, then $y(x, a) = \int_0^x e^{(\lambda+a)(x-\sigma)} f(\sigma)\,d\sigma$ and

$$y_n(x) = ((n-1)!)^{-1} \int_0^x (x-\sigma)^{n-1} e^{\lambda(x-\sigma)} f(\sigma)\,d\sigma.$$

If the q operators $L_1(x, D_x), \ldots, L_q(x, D_x)$ $(L_i \neq L_j)$ commute, then the above method can be extended to construct a solution $\tilde{y}(x)$ of the "iterated" nonhomogeneous equation $L_1^{m_1}(x, D_x) \cdots L_q^{m_q}(x, D_x)\tilde{y}(x) = f(x)$. To obtain $\tilde{y}(x)$, we solve the parameterized problem

$$\left(\prod_{n=1}^{q} (L_n(x, D_x) - a_n)\right) y(x, a_1, \ldots, a_q) = f(x), \tag{13}$$

and then evaluate the following expression

$$\left(\prod_{n=1}^{q}\left[(m_n-1)!\right]^{-1}\right)D_{a_1}^{m_1-1}\cdots D_{a_q}^{m_q-1}y(x,a_1,\ldots,a_q)|_{a_1=a_2=\cdots=a_q=0},$$

This method can also be applied to partial differential equations.

REFERENCES

1. M. A. Abramowitz and I. A. Stegun, eds., Handbook of Mathematical Functions, Nat. Bur. of Stds., Appl. Math. Series 55, Washington, D.C., 1964.
2. C. H. Edwards, and D. E. Penney, Calculus and Analytic Geometry, Prentice-Hall, Englewood Cliffs, 1982.
3. R. P. Feynman, Surely You're Joking, Mr. Feynman, W. W. Norton, New York, 1985.
4. P. Franklin, Methods of Advanced Calculus, McGraw-Hill, New York, 1944.
5. W. Fulks, Advanced Calculus, Wiley, New York, 1969.
6. D. V. Widder, Advanced Calculus, Prentice-Hall, Englewood Cliffs, 1961.
7. E. B. Wilson, Advanced Calculus, Ginn, Boston, 1912.
8. A. J. Zajta and S. K. Goel, Parametric integration techniques, *Mathematics Magazine*, 62 (1989) 318–322.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, JOHN A. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

An asterisk () after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before August 31, 1991 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY PROBLEM has already been solved in the literature, please tell the editors and include *copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3427. *Proposed by R. Padmanabhan, N. S. Mendelsohn, and B. Wolk, University of Manitoba, Winnipeg, Canada.*

Let S denote the set obtained by formally adjoining an element e to the set Q of rational numbers. On S define a binary operation $\circ$ as follows:

$$p \circ q = (3 + pq)/(p + q) \quad \text{if} \quad p \in Q, q \in Q, p \neq -q,$$
$$p \circ q = e \quad \text{if} \quad p \in Q, q \in Q, p = -q,$$
$$x \circ e = x \quad \text{for all} \quad x \in S.$$

(a) Prove that $(S, \circ)$ is an abelian group isomorphic to a subgroup of the multiplicative group of non-zero real numbers.

(b) If p is a positive rational number, put $p_1 = p$, $p_2 = p \circ p$, $p_3 = p \circ p \circ p, \ldots$. Show that $\lim_{n \to \infty} p_n$ exists and find the limit. For which values of p is the sequence $\{p_n\}_{n=1}^{\infty}$ monotonic?

E 3428. *Proposed by Artin B. Boghossian, Aramex Investments Inc., Willowdale, Ontario, Canada*

Let S be a non-empty interval on the real line. Let $f: S \to S$ be a continuous function having the property that for each $x \in S$ there exists a positive integer $n = n(x)$ with $f^n(x) = x$, where f^n denotes the n-th iterate of f. For given S characterize all such functions.

Part (a) was solved also by N. K. Artemiadis (Greece), G. Boros, D. Callan, S. Gaignoux (France), R. A. Groenveld, P. L. Hon (Canada), W. Janous (Austria), S. C. Kwok, L. Kuipers (Switzerland), K.-W. Lau (Hong Kong), I. E. Leonard (Canada), R. Lee, I. A. Sakmar (Turkey), V. Schindler (Germany), M. Saravi (Iran), P. Ševera & O. Šuch (Czechoslovakia), and the Lamar University Problem Solving Group, which also solved part (b). Two (incorrect) counterexamples were received.

Small Arcs

E 3324 [1989, 357]. *Proposed by Mark D. Meyerson and R. Bruce Richter, U.S. Naval Academy, Annapolis, MD.*

Suppose that a and b are two closed arcs (homeomorphs of $[0, 1]$) in the plane, each of which contains the origin O but not as an endpoint. Prove or disprove: (a) there are arbitrarily small neighborhoods N of O such that $N \cap a$ is connected. (b) there are arbitrarily small neighborhoods N of O such that both $N \cap a$ and $N \cap b$ are connected.

Solution by Richard Stong, Harvard University, Cambridge, MA. Statement (a) is true and statement (b) is false.

For (a), let $f\colon [0, 1] \to a$ be an arbitrary homeomorphism and let U be any open neighborhood of O small enough to miss the endpoints of a. Then $U \cap a$ is open in a, so $f^{-1}(U \cap a)$ is a disjoint union of open intervals. Let (s, t) be the open interval containing $f^{-1}(O)$. Then $N = U - (f[0, s]) \cup f([t, 1]))$ is an open neighborhood of O inside U such that $N \cap a = f((s, t))$ is connected.

For (b), let $\{a_i\}$ and $\{b_i\}$ be sequences converging to 0 such that $1 = a_1 > b_1 > a_2 > b_2 > \cdots > 0$. Let f be a continuous nonnegative function on $[-1, 1]$ with (1) $f(x) = 0$ for $x \leqslant 0$, (2) $f(a_i) = 0$ for all i, and (3) $f(b_i) > b_{i-2}$ for $i \geqslant 3$. Let $a = \{(x, y)\colon -1 \leqslant x \leqslant 1 \text{ and } y = f(x)\}$, and let $b = \{(x, y)\colon -1 \leqslant y \leqslant 1 \text{ and } x = f(y)\}$. Suppose N is a neighborhood of O such that $N \cap a$ and $N \cap b$ are connected. Then $(a_n, 0) \in N$ for some n. Hence the arc in a from $(a_n, 0)$ to $(b_n, f(b_n))$ is in N. This arc meets the arc in b from $(f(b_{n-2}), b_{n-2})$ to $(0, a_{n-1})$, because $f(b_i) > b_{i-2}$. Therefore, $(0, a_{n-1}) \in N$. Iterating this argument shows that $(a_n, 0)$ and $(0, a_n)$ belong to N for $n \geqslant 2$. Thus it is impossible for arbitrarily small neighborhoods of O to have a connected intersection with both a and b.

Solved also by A. W. Schurle, J. H. Steelman, and the proposers. One incorrect solution was received.

Must a Product of Derivatives Be a Derivative?

E 3329 [1989, 446]. *Proposed by Michel Balazard, Faculté des Sciences, Limoges, France.*

Suppose f and g are differentiable real-valued functions defined on $(-\infty, +\infty)$. Must there exist a differentiable real-valued function h defined on $(-\infty, +\infty)$ such that $h' = f'g'$?

Solution found independently by Michael W. Botsko, St. Vincent College, Latrobe, PA, Jean-Pierre Grivaux, Paris, and Marcin E. Kuczma, University of Warsaw, Poland. The answer is no. Let $\varphi(x) = \cos(x^{-1})$ if $x \neq 0$, $\varphi(0) = 0$. We will show that φ has a primitive, but that φ^2 does not. First, let $\psi(x) = x \sin(x^{-1})$ for $x \neq 0$ and $\psi(0) = 0$. Then ψ is continuous everywhere, and the function $x \cdot \psi$ is differentiable everywhere. The derivative of the latter is $2\psi - \varphi$, which thus has a primitive. Since ψ is continuous, φ must have some primitive f. If $\varphi^2 = h'$

everywhere, for some h, then

$$\frac{d}{dx}(h(x) - f(x/2)) = \varphi^2(x) - \tfrac{1}{2}\varphi(x/2)$$

$$= \begin{cases} \cos^2(1/x) - \frac{1}{2}\cos(2/x) = \frac{1}{2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

One can show directly that no differentiable function can have such a derivative, or one can appeal to the Darboux Theorem on intermediate values of derivatives. See, e.g., §129 of G. H. Hardy, *Pure Mathematics*, Cambridge University Press, or §21 of R. P. Boas, *A Primer of Real Functions*, Carus Mathematical Monographs No. 13.

Editorial comment. The solution above is essentially the same as an argument originally given by Witold Wilkosz, "Some properties of derivative functions," *Fund. Math.* 2(1921), 145–154. Several readers supplied that reference, as well as references to two articles by Andrew Bruckner: "Creating differentiability and destroying derivatives," this *Monthly* 85(1978), 554–562, and "Derivatives: Why they elude classification," *Math. Mag.* 49(1976), 5–11. Hugh Noland comments that in the latter article Bruckner cites Choquet's result: if h is continuous, then $h(f)$ is a derivative for every derivative f if and only if h is linear, and then argues "Suppose both f and f^2 are derivatives on $[\mathbf{0}, \mathbf{1}]$; then f^2, being non-negative, is the derivative of an increasing function F. But that makes f^2 integrable in the Lebesgue sense, and guarantees that f is also Lebesgue integrable. Thus, we need only take a derivative which is not Lebesgue integrable to furnish an example of a derivative whose square is not a derivative." Noland concludes by noting that $f(x) = x^{-1}\sin(x^{-2})$ for $x \neq 0$ and $f(0) = 0$ provides such an example; it is the derivative of

$$\int_0^x t\sin(t^{-2})\,dt - (x^2/2)\sin(x^{-2})$$

Solved also by A. A. Jagers (Netherlands), D. J. Johnson, I. Kozma (Israel), O. P. Lossers (Netherlands), X. Liu, K. Merryfield, J.-M. Monier (France), R. Mortini (Germany), H. Noland, A. Riese, R. R. Rogers, K. Schilling, N. Starr, J.-Y. Thibon (France), and the proposer.

A Characteristic of Low Degree Polynomials

E 3338 [1989, 641]. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Let s and t be given real numbers. Find all differentiable functions f on the real line that satisfy

$$f'(sx + ty) = \frac{f(y) - f(x)}{y - x}$$

for all real x, y with $x \neq y$.

Solution by Mordechai Falkowitz, Tel-Aviv, Israel. The only such functions are $f(x) = ax + b$, except that when $s = t = \frac{1}{2}$ any polynomial of degree 2 also works.

If $s + t \neq 0$, then setting $x = (z - t)/(s + t)$ and $y = (z + s)/(s + t)$ for arbitrary real z yields $f'(z) = f(y) - f(x)$, so that f' is differentiable. Since $s + t \neq 0$, we may assume $t \neq 0$ and differentiate the relationship $(y - x)f'(sx + ty) = f(y) - f(x)$ with respect to x to obtain

$$-f'(sx + ty) + (y - x)sf''(sx + ty) = -f'(x).$$

For the value $y = -(s/t)x$, we obtain $f'(x) = f'(0) + (s + t)(s/t)f''(0)x$ and conclude that f is a polynomial of degree at most 2. Testing $f(x) = ax^2 + \beta x + \gamma$ shows that α can be nonzero if and only if $s = t = \frac{1}{2}$. The computation for any $y \neq x$ is $2\alpha(sx + ty) + \beta = [\alpha(y^2 - x^2) + \beta(y - x)]/(y - x)$, or $2\alpha(sx + ty) = \alpha(x + y)$.

If $s + t = 0$, put $y = x + u$ for some $u \neq 0$. We have $f'(tu) = [f(x + u) - f(x)]/u$ for all x. Differentiating with respect to x yields $f'(x + u) - f'(x) = 0$. Setting $x = 0$ yields $f'(u) = f'(0)$. Since this holds for arbitrary u, f' is constant and the proof is complete.

Editorial comment. Many would-be solvers assumed without proof that in all cases of s, t, the function f must be twice differentiable everywhere.

John A. Baker proved the following generalization:

THEOREM. *If s, t are given real numbers, then the necessary and sufficient condition for f, g to satisfy $f(y) - f(x) = (y - x)g(sx + ty)$ for all x, y is as follows: If $s = t = 0$ or $s^2 \neq t^2$, then f is a polynomial of degree at most 1 and $g = f'$. If $s = t \neq 0$, then f is a polynomial of degree at most 2 and $g(x) = f'(x/2t)$. If $s = -t \neq 0$, then $f(x) = a + A(x)$ and $g(x) = A(x/t)/(x/t)$ for $x \neq 0$, where a is a real constant and A is an additive function (i.e., a function on the real line which satisfies $A(u + v) = A(u) + A(v)$ for all (u, v)).*

M. S. Jacobson & P. K. Sahoo provided references for published solutions of the problem $s = t = \frac{1}{2}$ (J. Aczel, "A mean value property of the derivative of quadratic polynomials—without mean values and derivatives," *Math. Mag.* 58(1985), 42–45, and S. Haruki, "A property of quadratic polynomials," this *Monthly* 86(1979), 577–579), and for a generalization of the problem (G. E. Cross and Pl. Kannappan, "A functional identity characterizing polynomials," *Aequat. Math.* 34(1987), 147–152).

Solved also by the proposer and 22 other readers. Several further solutions received contained errors, omissions, or hidden assumptions.

A Sequence of Polynomials Related to the Roots of Unity

E 3339 [1989, 641]. *Proposed by Antonio J. Duran, University of Seville, Spain.*

If k is a given positive integer, prove that we can find a polynomial g_k with rational coefficients and degree at most k such that

$$\sum_{j=1}^{n-1}{}' (e^{2\pi ij/n} - 1)^{-k} = g_k(n)$$

for any positive integer n. For example,

$$g_1(n) = -(n - 1)/2, \qquad g_2(n) = -(n - 1)(n - 5)/12,$$

$$g_3(n) = (n - 1)(n - 3)/8.$$

Solution by R. J. Chapman and R. W. K. Odoni, University of Exeter, UK. Consider the formal power series in t

$$f(n, t) = \sum_{k=1}^{\infty} g_k(n)\frac{t^k}{k}.$$

For the value $y = -(s/t)x$, we obtain $f'(x) = f'(0) + (s + t)(s/t)f''(0)x$ and conclude that f is a polynomial of degree at most 2. Testing $f(x) = ax^2 + \beta x + \gamma$ shows that α can be nonzero if and only if $s = t = \frac{1}{2}$. The computation for any $y \neq x$ is $2\alpha(sx + ty) + \beta = [\alpha(y^2 - x^2) + \beta(y - x)]/(y - x)$, or $2\alpha(sx + ty) = \alpha(x + y)$.

If $s + t = 0$, put $y = x + u$ for some $u \neq 0$. We have $f'(tu) = [f(x + u) - f(x)]/u$ for all x. Differentiating with respect to x yields $f'(x + u) - f'(x) = 0$. Setting $x = 0$ yields $f'(u) = f'(0)$. Since this holds for arbitrary u, f' is constant and the proof is complete.

Editorial comment. Many would-be solvers assumed without proof that in all cases of s, t, the function f must be twice differentiable everywhere.

John A. Baker proved the following generalization:

THEOREM. *If s, t are given real numbers, then the necessary and sufficient condition for f, g to satisfy $f(y) - f(x) = (y - x)g(sx + ty)$ for all x, y is as follows: If $s = t = 0$ or $s^2 \neq t^2$, then f is a polynomial of degree at most 1 and $g = f'$. If $s = t \neq 0$, then f is a polynomial of degree at most 2 and $g(x) = f'(x/2t)$. If $s = -t \neq 0$, then $f(x) = a + A(x)$ and $g(x) = A(x/t)/(x/t)$ for $x \neq 0$, where a is a real constant and A is an additive function (i.e., a function on the real line which satisfies $A(u + v) = A(u) + A(v)$ for all (u, v)).*

M. S. Jacobson & P. K. Sahoo provided references for published solutions of the problem $s = t = \frac{1}{2}$ (J. Aczel, "A mean value property of the derivative of quadratic polynomials—without mean values and derivatives," *Math. Mag.* 58(1985), 42–45, and S. Haruki, "A property of quadratic polynomials," this *Monthly* 86(1979), 577–579), and for a generalization of the problem (G. E. Cross and Pl. Kannappan, "A functional identity characterizing polynomials," *Aequat. Math.* 34(1987), 147–152).

Solved also by the proposer and 22 other readers. Several further solutions received contained errors, omissions, or hidden assumptions.

A Sequence of Polynomials Related to the Roots of Unity

E 3339 [1989, 641]. *Proposed by Antonio J. Duran, University of Seville, Spain.*

If k is a given positive integer, prove that we can find a polynomial g_k with rational coefficients and degree at most k such that

$$\sum_{j=1}^{n-1} (e^{2\pi ij/n} - 1)^{-k} = g_k(n)$$

for any positive integer n. For example,

$$g_1(n) = -(n - 1)/2, \qquad g_2(n) = -(n - 1)(n - 5)/12,$$

$$g_3(n) = (n - 1)(n - 3)/8.$$

Solution by R. J. Chapman and R. W. K. Odoni, University of Exeter, UK. Consider the formal power series in t

$$f(n, t) = \sum_{k=1}^{\infty} g_k(n) \frac{t^k}{k}.$$

Richard Guy provided a table computed by H. K. Farahat. The next values of $g_k(n)$ are

$$g_4(n) = (n-1)(n^3+n^2-109n+251)/(2^4\cdot 3^2\cdot 5)$$

$$g_5(n) = -(n-1)(n-5)(n^2+6n-19)/(2^5\cdot 3^2)$$

$$g_6(n) = -(n-1)(2n^5+2n^4-355n^3-355n^2+11153n-19087)/(2^6\cdot 3^3\cdot 5\cdot 7)$$

$$g_7(n) = (n-1)(n-7)(2n^4+16n^3-33n^2-376n+751)/(2^7\cdot 3^3\cdot 5)$$

$$g_8(n) = (n-1)(3n^7+3n^6-917n^5-917n^4+39697n^3+39697n^2-744383n+1070017)/(2^8\cdot 3^4\cdot 5^2\cdot 7)$$

$$g_9(n) = -(n-1)(n-3)(n-9)(n^5+13n^4+10n^3-350n^2-851n+2857)/(2^9\cdot 3\cdot 5^2\cdot 7).$$

Most solvers noted that the numbers $(e^{2\pi ij/n}-1)^{-1}$ $(1 \leqslant j \leqslant n-1)$ are the roots of $(w+1)^n - w^n = 0$ and so their k-th elementary symmetric function is $(-1)^k \binom{n-1}{k}/(k+1)$ for $k = 1, 2, \ldots$. The power sums $g_k(n)$ can then be calculated recursively by using Newton's formulas.

Solved also by S. J. Bang (Korea), D. Callan, J. Ferrer (Spain), S. M. Gagola, Jr., I. Gessel, A. A. Jagers (The Netherlands), W. Janous (Austria), H. Ki (Korea), E. Levine, O. P. Lossers (The Netherlands), J. McHugh, H. C. Morris, J. H. Nieto (Venezuela), A. Nijenhuis, H. C. Wei (China), and the proposer.

A Transcendental Sum

E 3353 [1989, 838]. *Proposed by H. G. Diamond, University of Illinois, Urbana, IL.*

For each positive integer n, let $G(n)$ be the number of digits in the decimal expansion of 2^n. Is it true that

$$\sum_{n=1}^{\infty} \frac{G(n)}{2^n} = \frac{1169}{1023}?$$

Solution I by Kee-Wai Lau, Hong Kong. We show that the desired sum S exceeds $1169/1023$. Using $\log x$ for $\log_{10} x$, we have $S = \sum_{n=1}^{\infty} \lfloor 1 + n\log 2 \rfloor / 2^n$, or

$$S = 1 + \sum_{n=1}^{10} \frac{\lfloor n\log 2 \rfloor}{2^n} + \sum_{n=1}^{\infty} \frac{\lfloor (n+10)\log 2 \rfloor}{2^{n+10}}$$

$$= \frac{1167}{1024} + \frac{1}{1024} \sum_{n=1}^{\infty} \frac{\lfloor (n+10)\log 2 \rfloor}{2^n}.$$

Since $\lfloor (n+10)\log 2 \rfloor \geqslant \lfloor n(\log 2) + 3 \rfloor$, with strict inequality at $n = 93$, we have

$$S > \frac{1167}{1024} + \frac{1}{1024} \sum_{n=1}^{\infty} \frac{\lfloor n(\log 2) + 3 \rfloor}{2^n} = \frac{1167}{1024} + \frac{1}{1024}(S+2),$$

which implies $S > 1169/1023$.

Solution II by Reiner Martin, Karlsruhe, Germany. If n_k is the least positive integer with $G(n_k) = k$, then $1 = n_1 < n_2 < \cdots$ and $S = \sum_{k=1}^{\infty} \sum_{n=n_k}^{\infty} 2^{-n} = \sum_{k=1}^{\infty} 2^{1-n_k}$. Since $2^{102} < 10^{31} < 2^{103} < 10^{32}$, we have $n_{32} = 103$. Hence the 102nd

digit in the binary expansion of S is 1, but the 102nd digit in the binary expansion of

$$1169/1023 = 1 + 146\sum_{n=1}^{\infty} 2^{-10n} = 1 + \sum_{n=1}^{\infty}(2^{7-10n} + 2^{4-10n} + 2^{1-10n})$$

is zero.

Solution III by B. M. M. de Weger, University of Twente, Enschede, The Netherlands. William W. Adams and J. L. Davison, in "A remarkable class of continued fractions," *Proc. Amer. Math. Soc.* 65(1977) 194–198 showed that if α is irrational, then $T(\alpha) = \sum_{k=1}^{\infty} 2^{-\lfloor k\alpha \rfloor}$ is transcendental. Now it is easy to see that (in the notation of Solution II)

$$n_k = 1 + \lfloor (k-1)\log 10/\log 2 \rfloor .$$

Thus $S = 1 + T(\log 10/\log 2)$. Since $\log 10/\log 2$ is irrational by the unique factorization theorem, S is transcendental and cannot equal the rational number 1169/1023.

Editorial Comment. Adams and Davison proved their result by showing that the partial quotients in the continued fraction expansion of $T(\alpha)$ increase more rapidly than is possible for an algebraic number (by Roth's Theorem). In the case of S the first few partial quotients are

$$1, 7, 146, \lfloor 2^{93}/1023 \rfloor = 2^3(2^{90} - 1)/(2^{10} - 1), \ldots .$$

Thus, $1 + 1/(7 + 1/146) = 1169/1023$ is a very good rational approximation to S.

Michael Filaseta solved the problem by using Roth's Theorem to prove that if α is irrational, then $\sum_{n=1}^{\infty} \lfloor n\alpha \rfloor 2^{-n}$ is transcendental. Now, if α is an irrational number between 0 and 1, Doug Bowman remarked that

$$(1-x)\sum_{n=1}^{\infty} \lfloor n\alpha \rfloor x^n = x\sum_{n=1}^{\infty} x^{\lfloor n/\alpha \rfloor} \qquad (|x| < 1).$$

Thus Filaseta's result is equivalent to the theorem of Adams and Davison quoted in Solution III.

The smallest positive integer N such that

$$\sum_{n=1}^{N} G(n) \cdot 2^{-n} > 1169/1023 \text{ is } N = 109.$$

Cf. the solution of Problem 6609 in this issue.

Thirty-three correct solutions and three incorrect solutions were received.

ADVANCED PROBLEMS

6652. *Proposed by D. M. Bloom, Brooklyn College of the City University of New York.*

For x a positive integer put

$$E(x) = \sum_{0 \leq i \leq x} \frac{(-1)^i}{i!}(x-i)^i e^{x-i}.$$

Evaluate

$$\lim_{x\to\infty}\{E(x) - 2x\}.$$

6653. *Proposed by D. K. Lee, University of Ulsan, Korea, and B. A. Murray, University of Newcastle-upon-Tyne, UK.*

For non-negative real a evaluate the integral

$$I(a) = \int_0^{\pi} \psi(\theta)\sin\theta\, d\theta,$$

where $\psi(\theta) = \arccos(\{a - \cos\theta\}/\sqrt{1 + a^2 - 2a\cos\theta})$, $0 \leqslant \psi(\theta) \leqslant \pi$.

6654. *Proposed by W. O. Egerland and C. E. Hansen, Aberdeen Proving Ground, Aberdeen, MD.*

Suppose ω is real, n is a positive integer greater than 1, and $a_1, a_2, \ldots, a_n$ are complex numbers with $|a_k| < 1$ for $k = 1, 2, \ldots, n$. Prove that the equation

$$e^{i\omega}(z - a_1)(z - a_2)\cdots(z - a_n) = z(1 - \bar{a}_1 z)(1 - \bar{a}_2 z)\cdots(1 - \bar{a}_n z)$$

has at least $n - 1$ roots on the unit circle.

SOLUTIONS OF ADVANCED PROBLEMS

Successive Divisor Sums

6592 [1989, 165]. *Proposed by Dietrich Haase, Free University of Berlin, and Adolf Hildebrand, University of Illinois at Urbana-Champaign.*

For real k let $\sigma_k(n)$ denote the sum of the k-th powers of the (positive) divisors of the natural number n. Show that there exists a number k_0 satisfying $1 < k_0 < 2$ such that

(i) if $k > k_0$, then $\{\sigma_k(n + 1) - \sigma_k(n)\}(-1)^{n+1} > 0$ for all sufficiently large n,

(ii) if $0 < k < k_0$, then $\{\sigma_k(n + 1) - \sigma_k(n)\}(-1)^{n+1}$ changes sign infinitely often. (Cf. Problem 6555.)

Editorial remark. Through an unfortunate editorial blunder, Hildebrand was listed as the sole proposer in the February, 1989 issue. The assertion of part (i) is due to Haase; the assertion of part (ii) is due to Hildebrand.

Composite solution, based on the solutions of the proposers and Eugene Levine, Adelphi University, Garden City, NY. Put

$$S_k(n) = (-1)^{n+1}\{\sigma_k(n + 1) - \sigma_k(n)\}.$$

Let $f\colon (1, \infty) \to (0, \infty)$ be the decreasing continuous function defined by

$$f(x) = \sum_{t=1}^{\infty} (t + 1/2)^{-x} \qquad (x > 1).$$

In terms of the Riemann zeta-function we have $f(x) = (2^x - 1)\zeta(x) - 2^x$. Let k_0 be the unique real number greater than 1 such that $f(k_0) = 1$, that is, $k_0 = 1.9401\ldots$. With this choice of k_0 we strengthen (i) and (ii) by proving the following assertions.

Evaluate

$$\lim_{x\to\infty}\{E(x) - 2x\}.$$

6653. *Proposed by D. K. Lee, University of Ulsan, Korea, and B. A. Murray, University of Newcastle-upon-Tyne, UK.*

For non-negative real a evaluate the integral

$$I(a) = \int_0^{\pi} \psi(\theta)\sin\theta\, d\theta,$$

where $\psi(\theta) = \arccos(\{a - \cos\theta\}/\sqrt{1 + a^2 - 2a\cos\theta})$, $0 \leqslant \psi(\theta) \leqslant \pi$.

6654. *Proposed by W. O. Egerland and C. E. Hansen, Aberdeen Proving Ground, Aberdeen, MD.*

Suppose ω is real, n is a positive integer greater than 1, and $a_1, a_2, \ldots, a_n$ are complex numbers with $|a_k| < 1$ for $k = 1, 2, \ldots, n$. Prove that the equation

$$e^{i\omega}(z - a_1)(z - a_2)\cdots(z - a_n) = z(1 - \bar{a}_1 z)(1 - \bar{a}_2 z)\cdots(1 - \bar{a}_n z)$$

has at least $n - 1$ roots on the unit circle.

SOLUTIONS OF ADVANCED PROBLEMS

Successive Divisor Sums

6592 [1989, 165]. *Proposed by Dietrich Haase, Free University of Berlin, and Adolf Hildebrand, University of Illinois at Urbana-Champaign.*

For real k let $\sigma_k(n)$ denote the sum of the k-th powers of the (positive) divisors of the natural number n. Show that there exists a number k_0 satisfying $1 < k_0 < 2$ such that

(i) if $k > k_0$, then $\{\sigma_k(n + 1) - \sigma_k(n)\}(-1)^{n+1} > 0$ for all sufficiently large n,

(ii) if $0 < k < k_0$, then $\{\sigma_k(n + 1) - \sigma_k(n)\}(-1)^{n+1}$ changes sign infinitely often. (Cf. Problem 6555.)

Editorial remark. Through an unfortunate editorial blunder, Hildebrand was listed as the sole proposer in the February, 1989 issue. The assertion of part (i) is due to Haase; the assertion of part (ii) is due to Hildebrand.

Composite solution, based on the solutions of the proposers and Eugene Levine, Adelphi University, Garden City, NY. Put

$$S_k(n) = (-1)^{n+1}\{\sigma_k(n + 1) - \sigma_k(n)\}.$$

Let $f\colon (1, \infty) \to (0, \infty)$ be the decreasing continuous function defined by

$$f(x) = \sum_{t=1}^{\infty} (t + 1/2)^{-x} \qquad (x > 1).$$

In terms of the Riemann zeta-function we have $f(x) = (2^x - 1)\zeta(x) - 2^x$. Let k_0 be the unique real number greater than 1 such that $f(k_0) = 1$, that is, $k_0 = 1.9401\ldots$. With this choice of k_0 we strengthen (i) and (ii) by proving the following assertions.

I. If $|k| \geqslant k_0$, then $S_k(n) > 0$ for all sufficiently large positive integers n.

II. If $|k| < k_0$, then

(a) $S_k(n) > 0$ for infinitely many odd values of n,

(b) $S_k(n) > 0$ for infinitely many even values of n,

(c) $S_k(n) < 0$ for infinitely many odd values of n,

(d) $S_k(n) < 0$ for infinitely many even values of n.

(The method of solution of 6555 [1987, 800; 1990, 351] would probably make it possible to replace (a), (b), (c), (d) by positive density assertions.)

The extension of (i) and (ii) to non-positive values of k is straightforward, but the extension of (i) to the case $|k| = k_0$ requires the following lemma.

LEMMA. *If n is a sufficiently large positive integer, there exists a prime number q not exceeding $n^{1/2}$ and not dividing $n + 1$.*

Proof of Lemma. We use Bertrand's postulate (Niven and Zuckerman, *An Introduction to the Theory of Numbers*, §8.3). Suppose $n \geqslant 256$ and $r = \lfloor n^{1/2}/8 \rfloor$. Bertrand's postulate asserts the existence of primes p_1, p_2, p_3 in the intervals $(r, 2r)$, $(2r, 4r)$, and $(4r, 8r)$ respectively. Then

$$p_1 p_2 p_3 \geqslant \lfloor n^{1/2}/8 \rfloor^3 > n + 1$$

for $n > 10^6$, and so at least one of the primes p_1, p_2, p_3 does not divide $n + 1$. Thus the lemma is proved.

Proof of I. We begin by noting that if $\theta_k = 1 + 2^{-k}$ and if

$$\lambda_k = \sum_{t=0}^{\infty} (2t + 1)^{-k} = 1 + 2^{-k} f(k) \quad \text{for } k > 1,$$

then $\theta_k > \lambda_k$ for $k > k_0$, $\theta_k = \lambda_k$ for $k = k_0$, and $\theta_k < \lambda_k$ for $1 < k < k_0$. Further we have

$$\sigma_k(n)/n^k = \sigma_{-k}(n) \geqslant \theta_k \quad \text{for even } n,$$

$$\sigma_k(n)/n^k = \sigma_{-k}(n) < \lambda_k \quad \text{for } n \text{ odd}, k > 1,$$

$$\sigma_k(n)/n^k = \sigma_{-k}(n) < \lambda_k - n^{-k} \quad \text{for } n \text{ odd}, k > 1,$$

where m is any odd integer not dividing n.

If $k \geqslant k_0$ and n is odd, then

$$\frac{\sigma_k(n+1)}{\sigma_k(n)} = \frac{(n+1)^k \sigma_{-k}(n+1)}{n^k \sigma_{-k}(n)} > \frac{\sigma_{-k}(n+1)}{\sigma_{-k}(n)} > \frac{\theta_k}{\lambda_k} \geqslant 1.$$

Thus $S_k(n) > 0$ for $|k| \geqslant k_0$ and n odd.

If $k > k_0$ and n is even, we have

$$\frac{\sigma_{-k}(n)}{\sigma_{-k}(n+1)} = \frac{(n+1)^k \sigma_k(n)}{n^k \sigma_k(n+1)} > \frac{\sigma_k(n)}{\sigma_k(n+1)} > \frac{\theta_k n^k}{\lambda_k (n+1)^k} > 1,$$

providing n is sufficiently large. Thus $S_k(n) > 0$ for $|k| > k_0$ and n a sufficiently large even number.

provided p is sufficiently large. Also

$$\frac{\sigma_k(2p)}{\sigma_k(2p-1)} = \frac{(2p)^k}{(2p-1)^k}\frac{\sigma_{-k}(2p)}{\sigma_{-k}(2p-1)} < 1$$

if p is sufficiently large. Thus if $0 < |k| < k_0$, we have $S_k(2p-1) < 0$ for any sufficiently large prime p satisfying the congruence $p \equiv (b+1)/2 \pmod{b}$. Hence assertion (c) holds for $0 < |k| < k_0$.

If $0 < k < k_0$, we may again choose N so large that $\sum_{t=1}^{N}(t+\frac{1}{2})^{-k} = 1 + c$, where $c > 0$. By Dirichlet's theorem there are infinitely many primes p such that $2p+1$ is divisible by $b = 3 \cdot 5 \cdot 7 \ldots (2N+1)$, i.e., such that $p \equiv (b-1)/2 \pmod{b}$. Then

$$\begin{aligned}\frac{\sigma_k(2p+1)}{\sigma_k(2p)} &= \frac{(2p+1)^k\sigma_{-k}(2p+1)}{(2p)^k\sigma_{-k}(2p)} > \frac{\sigma_{-k}(2p+1)}{\sigma_{-k}(2p)} \\ &\geqslant \frac{1+3^{-k}+5^{-k}+\cdots+(2N+1)^{-k}}{1+2^{-k}+p^{-k}+(2p)^{-k}} \\ &= \frac{1+2^{-k}+2^{-k}c}{(1+2^{-k})(1+p^{-k})} > 1,\end{aligned}$$

provided p is sufficiently large. Thus, if $0 < |k| < k_0$, we have $S_k(2p) < 0$ for any sufficiently large prime p satisfying the congruence $p \equiv (b-1)/2 \pmod{b}$. Hence assertion (d) holds for $0 < |k| < k_0$.

Solved also by Paul Erdös (Hungary) and Joachim Herzog (Germany).

Circles Passing Through an Odd Number of Lattice Points

6605 [1989, 533]. *Proposed by E. Ehrhart, Université de Strasbourg, France.*

We use the term "lattice point" to mean a point with integer coordinates.

If k is a positive integer, Schinzel [*Enseignement Math.* (2) 4 (1958) 71–72] proved that the circle

$$\left(X - \frac{1}{3}\right)^2 + Y^2 = (5^k/3)^2$$

passes through exactly $2k+1$ lattice points; clearly the two coordinates of any one of these $2k+1$ lattice points are of like parity. Thus by making the substitution $X = x + y$, $Y = x - y$ we see that the smaller circle

$$\left(x - \frac{1}{6}\right)^2 + \left(y - \frac{1}{6}\right)^2 = (5^k\sqrt{2}/6)^2 \qquad (*)$$

also passes through exactly $2k+1$ lattice points.

(i) Show that when $k = 1$ no circle smaller than $(*)$ passes through exactly three lattice points.

(ii) Show that when $k = 2$ no circle smaller than $(*)$ passes through exactly five lattice points.

(iii) Show that when $2k+1$ is composite there is a circle smaller than $(*)$ which passes through exactly $2k+1$ lattice points.

Solution by David Borwein, University of Western Ontario, London, Ontario. For (i) it is readily seen that the smallest circle passing through exactly three lattice

points must circumscribe a triangle congruent to $(0, -1)$, $(-1, 0)$, $(1, 1)$ and so have radius $5\sqrt{2}/6$. Any smaller circle passes through either an even number of lattice points or no lattice points. We do not have a simple deductive argument for the truth of (ii), but a computer search failed to yield any smaller circle passing through exactly five lattice points. For (iii), suppose $2k + 1 = (2r + 1)(2s + 1)$ where r and s are positive integers with $r \leqslant s$. By standard number theory the equation

$$u^2 + v^2 = 2 \cdot 5^{2r}13^{2s}$$

has exactly $4(2r + 1)(2s + 1) = 4(2k + 1)$ integer solutions. Further, since

$$2 \cdot 5^{2r}13^{2s} \equiv 2 \pmod 6,$$

all solutions must be of the form $u \equiv \pm 1 \pmod 6$, $v \equiv \pm 1 \pmod 6$; all other possibilities are easily excluded. Hence all solutions have the form $u = \pm(6x - 1)$, $v = \pm(6y - 1)$ where x and y are integers. Therefore the equation

$$(6x - 1)^2 + (6y - 1)^2 = 2 \cdot 5^{2r}13^{2s}$$

has exactly $2k + 1$ integer solutions, and so the circle

$$\left(x - \frac{1}{6}\right)^2 + \left(y - \frac{1}{6}\right)^2 = \left(5^r13^s\sqrt{2}/6\right)^2$$

passes through exactly $2k + 1$ lattice points. But $5^r13^s < 5^k$ since $13^s < 5^{k-r} = 5^{s(1+2r)}$, and so this circle has smaller radius than $(*)$. This completes the proof; we note that it is straightforward to show, along the lines of the argument used to establish (iii), that when $2k + 1$ is prime $(*)$ is the smallest circle with center $(1/6, 1/6)$ that passes through exactly $2k + 1$ lattice points.

Solved also by the proposer.

An Application of the Artin-Schreier Theorem

6606 [1989, 652]. *Proposed by Daniel B. Shapiro, Ohio State University, Columbus.*

Suppose k is a field containing an element i with $i^2 = -1$ and suppose $\tilde{k}$ is the algebraic closure of k. Suppose F and K are fields with $k \subseteq F \subseteq K$ and $[K : F]$ finite.

(1) If $\tilde{k} \subseteq K$, prove that in fact $\tilde{k} \subseteq F$.

(2) More generally, if L is an algebraically closed field with $k \subseteq L \subseteq K$, does it follow that $L \subseteq F$?

Solution by I. Martin Isaacs, University of Wisconsin, Madison. We shall prove (1) and then construct for (2) an example in which L is not contained in F. For (1) we may assume that $k = \tilde{k} \cap F$. Since $i \in k$, it follows via the Artin-Schreier theorem that if $\tilde{k}$ properly contains k, then there exists $\alpha \in \tilde{k}$ with minimal polynomial g (over k) of arbitrarily large degree. The minimal polynomial f of α over F, however, divides this polynomial, and so all of its roots (in an appropriate extension field) are algebraic over k and so lie in $\tilde{k}$. It follows that all the coefficients of f lie in k and so f divides g in the ring $k[x]$. As g is irreducible, we deduce that $f = g$ and so $\deg(g) = \deg(f) \leqslant [K : F]$. This contradicts the fact that $\deg(g)$ can be made arbitrarily large. We deduce that $\tilde{k} = k$.

For (2), suppose k is any algebraically closed field of characteristic zero and let $L \supseteq k$ be an algebraic closure of a simple transcendental extension of k. Write $R = L \otimes_k L$ so that R is a k-algebra. Note that R has subrings $L_1 = L \otimes 1$ and

$L_2 = 1 \otimes L$, each isomorphic to the algebraically closed field L, and also R has an automorphism σ mapping $x \otimes y$ to $y \otimes x$, so that σ has order 2 and interchanges L_1 and L_2. Since k is algebraically closed in L and has characteristic zero, Theorem 4.5.2 of Nagata's book *Field Theory* guarantees that R is an integral domain. Let K be its quotient field. Then the unique extension of σ to K has order 2. If F is the fixed field of this extension, then $k \subseteq F \subseteq K$ and F does not contain the algebraically closed field L although $|K : F| = 2$.

Editorial comment. A good reference for the Artin-Schreier theorem is the last section of Volume 3 of Nathan Jacobson's *Lectures in Abstract Algebra*, Van Nostrand, Princeton, 1964 or Section 12 of Irving Kaplansky's *Fields and Rings*, University of Chicago Press, 1972. In a paper of S. Abhyankar, "On the compositum of algebraically closed subfields," *Proc. Amer. Math. Soc.* 7(1956) 905–907, it is shown that (2) is true when K is finitely generated over k. In fact, $L \subseteq \tilde{k}$ in that case.

Solved also by R. J. Chapman & R. W. K. Odoni (United Kingdom) and by the proposer. Two incorrect solutions were received.

Polynomials Related to the Laguerre Polynomials

6607 [1989, 652]. *Proposed by Ira Gessel, Brandeis University, Waltham, MA.*

The Laguerre polynomials are defined by

$$L_n^{(\alpha)}(x) = (-1)^n \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{k} \frac{x^{n-k}}{(n-k)!},$$

where α is an arbitrary parameter and n is a non-negative integer. Show that for each non-negative integer k there exists a polynomial q_k of degree $\lfloor k/2 \rfloor$ depending only on k such that

$$L_n^{(\alpha)}(x+n+\alpha) = (-1)^n \sum_{k=0}^{n} q_k(n+\alpha) \frac{x^{n-k}}{(n-k)!}.$$

For example $q_0(u) = 1$, $q_1(u) = 0$, $q_2(u) = -u/2$, $q_3(u) = -u/3$, $q_4(u) = u^2/8 - u/4$.

Solution by Mourad E. H. Ismail, University of South Florida, Tampa. Replace x by $x + n + \alpha$ in the definition of a Laguerre polynomial and apply the binomial theorem. We obtain

$$L_n^{(\alpha)}(x+n+\alpha) = \sum_{k=0}^{n} \binom{n+\alpha}{k} (-1)^{n+k} \sum_{j=0}^{n-k} \frac{(\alpha+n)^j x^{n-k-j}}{j!(n-k-j)!}$$

$$= (-1)^n \sum_{m=0}^{n} q_m(n+\alpha) \frac{x^{n-m}}{(n-m)!},$$

where

$$q_m(u) = \sum_{k=0}^{m} \binom{u}{k} \frac{(-1)^k u^{m-k}}{(m-k)!}.$$

Clearly $q_m(u)$ depends only on m. Since $q_m(u)$ is a Cauchy product, we see that

$$\sum_{m=0}^{\infty} q_m(u)t^m = (1-t)^u e^{ut} = \exp\{u[t + \ln(1-t)]\}$$

$$= \exp\left[-ut^2 \sum_{n=0}^{\infty} \frac{t^n}{n+2}\right].$$

By expanding the right-hand side we find that $q_m(u)$ has precise degree $\lfloor m/2 \rfloor$ and that the coefficient of $u^{\lfloor m/2 \rfloor}$ is

$$(-1/2)^{\lfloor m/2 \rfloor}/(\lfloor m/2 \rfloor)!.$$

Solved also by Seung-Jin Bang (South Korea), David Callan, Michael Golomb, Bruno Haible (Germany), A. A. Jagers (The Netherlands), and the proposer.

A Rational Sum

6609 [1989, 743]. *Proposed by Doug Bowman and Tad White, University of California at Los Angeles.*

For each positive integer n let $g(n)$ be the number of digits greater than 4 in the decimal expansion of 2^n. Is it true that

$$\sum_{n=1}^{\infty} \frac{g(n)}{2^n} = \frac{2}{9}?$$

Editorial remark. Exactly the same result was obtained previously by Eugene Levine, who published it in the *College Mathematics Journal* as Problem 386 [1988, 448; 1990, 151]. Levine also generalized the result to even bases $b > 2$, with sum $2/(b-1)$ rather than $2/9$. There were eight CMJ solvers, one of whom (Dave Ohlsen) also proposed a generalization; among the nineteen AMM solvers several proposed generalizations (not all different).

Bowman and White actually provided the editors with a variety of related questions, only one of which duplicated Levine's problem. With hindsight the following would have been a better choice:

Define the Fibonacci sequence by $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}$ for $n \geqslant 3$. For each positive integer n let $f(n)$ be the sum of the Fibonacci numbers F_i over all integers i such that the ith digit of 2^n (counting from the right) exceeds 4. Show that

$$\sum_{n=1}^{\infty} \frac{f(n)}{2^n} = \frac{20}{89}. \qquad (*)$$

Solution by Doug Bowman and Tad White (of both the proposed problem and ().* Let $r_0, r_1, r_2, \ldots,$ be a sequence of integers with $r_n > 1$, and define another sequence a_n inductively by $a_0 = 1$ and $a_{n+1} = a_n r_n$. Let the expansion in base b of a_n be

$$a_n = \sum_{i \geqslant 0} d_{n,i} b^i.$$

Clearly $q_m(u)$ depends only on m. Since $q_m(u)$ is a Cauchy product, we see that

$$\sum_{m=0}^{\infty} q_m(u)t^m = (1-t)^u e^{ut} = \exp\{u[t + \ln(1-t)]\}$$

$$= \exp\left[-ut^2 \sum_{n=0}^{\infty} \frac{t^n}{n+2}\right].$$

By expanding the right-hand side we find that $q_m(u)$ has precise degree $\lfloor m/2 \rfloor$ and that the coefficient of $u^{\lfloor m/2 \rfloor}$ is

$$(-1/2)^{\lfloor m/2 \rfloor}/(\lfloor m/2 \rfloor)!.$$

Solved also by Seung-Jin Bang (South Korea), David Callan, Michael Golomb, Bruno Haible (Germany), A. A. Jagers (The Netherlands), and the proposer.

A Rational Sum

6609 [1989, 743]. *Proposed by Doug Bowman and Tad White, University of California at Los Angeles.*

For each positive integer n let $g(n)$ be the number of digits greater than 4 in the decimal expansion of 2^n. Is it true that

$$\sum_{n=1}^{\infty} \frac{g(n)}{2^n} = \frac{2}{9}?$$

Editorial remark. Exactly the same result was obtained previously by Eugene Levine, who published it in the *College Mathematics Journal* as Problem 386 [1988, 448; 1990, 151]. Levine also generalized the result to even bases $b > 2$, with sum $2/(b-1)$ rather than $2/9$. There were eight CMJ solvers, one of whom (Dave Ohlsen) also proposed a generalization; among the nineteen AMM solvers several proposed generalizations (not all different).

Bowman and White actually provided the editors with a variety of related questions, only one of which duplicated Levine's problem. With hindsight the following would have been a better choice:

Define the Fibonacci sequence by $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}$ for $n \geqslant 3$. For each positive integer n let $f(n)$ be the sum of the Fibonacci numbers F_i over all integers i such that the ith digit of 2^n (counting from the right) exceeds 4. Show that

$$\sum_{n=1}^{\infty} \frac{f(n)}{2^n} = \frac{20}{89}. \qquad (*)$$

Solution by Doug Bowman and Tad White (of both the proposed problem and ().* Let $r_0, r_1, r_2, \ldots,$ be a sequence of integers with $r_n > 1$, and define another sequence a_n inductively by $a_0 = 1$ and $a_{n+1} = a_n r_n$. Let the expansion in base b of a_n be

$$a_n = \sum_{i \geqslant 0} d_{n,i} b^i.$$

The computation of the expansion in base b of a_{n+1} may be displayed as follows:

$$\begin{array}{ccccc} \cdots c_{n,2} & & c_{n,1} & & c_{n,0} \\ \cdots d_{n,2} & & d_{n,1} & & d_{n,0} \\ & & & & \times r_n \\ \hline \cdots d_{n+1,2} & & d_{n+1,1} & & d_{n+1,0} \end{array}$$

Here the $c_{n,i}$ denote "carries," $c_{n,0} = 0$ for each n, and $0 \leqslant d_{n,i} < b$ for all n and i. Note that if $r_n < b$, the base b representation of r_n is a 1-digit number, and our ordinary elementary-school definition of "carry" is valid. If $r_n \geqslant b$, we multiply nonetheless as the diagram indicates; the multiplication is distributed over the base b expansion of a_n, but not over the expansion of r_n. In other words, the "next digit" and "next carry" are given by the equation

$$r_n d_{n,i} + c_{n,i} = d_{n+1,i} + bc_{n,i+1} \qquad (**)$$

for all $n \geqslant 0$ and $i \geqslant 0$. Now define, for z real or complex, the function

$$g(n; z) = \sum_{i \geqslant 0} c_{n,i} z^i.$$

PROPOSITION. *If* $|z| < b$, *then*

$$\sum_{n=0}^{\infty} \frac{g(n; z)}{a_{n+1}} = \frac{z}{b - z}.$$

Proof. Multiply the "next digit" equation $(**)$ by z^{i+1}/a_{n+1} and collect c's and d's to obtain

$$\frac{1}{a_{n+1}}\left(bz^{i+1}c_{n,i+1} - z \cdot z^i c_{n,i}\right) = z\left(\frac{z^i d_{n,i}}{a_n} - \frac{z^i d_{n+1,i}}{a_{n+1}}\right).$$

Since $c_{n,0} = 0$, summation on i yields

$$(b - z)\frac{g(n; z)}{a_{n+1}} = z \sum_{i \geqslant 0}\left(\frac{z^i d_{n,i}}{a_n} - \frac{z^i d_{n+1,i}}{a_{n+1}}\right).$$

Hence

$$\begin{aligned} \sum_{0 \leqslant n \leqslant N} (b - z)\frac{g(n; z)}{a_{n+1}} &= z \sum_{i \geqslant 0} z^i \sum_{0 \leqslant n \leqslant N}\left(\frac{d_{n,i}}{a_n} - \frac{d_{n+1,i}}{a_{n+1}}\right) \\ &= z \sum_{i \geqslant 0} \frac{z^i d_{0,i}}{a_0} - z \sum_{i \geqslant 0} \frac{z^i d_{N+1,i}}{a_{N+1}}. \end{aligned}$$

Since $d_{0,0} = 1$ and $d_{0,i} = 0$ for $i \geqslant 1$ (recall $a_0 = 1$), this says that

$$\sum_{0 \leqslant n \leqslant N} \frac{g(n; z)}{a_{n+1}} = \frac{z}{b - z}\left(1 - \frac{(a_{N+1} : z)}{a_{N+1}}\right),$$

where

$$(a_{N+1} : z) = \sum_{i \geqslant 0} d_{N+1,i} z^i,$$

i.e., $(a_{N+1} : z)$ is the result of writing a_{N+1} in base b and then replacing b by the (smaller in modulus) number z. It is easy to see that $(a_{N+1} : z)/a_{N+1} \to 0$ as $N \to \infty$, and the result follows.

For Levine's result take $a_n = 2^n$, $z = 1$, and b an even number greater than 2. The only possible carries are 1's: They occur whenever a digit of a_n exceeds $b/2 - 1$. As for $(*)$ note that $F_n = (\alpha^n - \beta^n)/(\alpha - \beta)$ where α and β are the roots of $z^2 - z - 1 = 0$. Take as before $a_n = 2^n$. The only carries in the computation of $2 \cdot 2^n$ occur when a digit of 2^n exceeds $b/2 - 1$, so that

$$f(n) = \frac{g(n;\alpha) - g(n;\beta)}{\alpha - \beta}.$$

The proposition yields

$$\sum_{n \geqslant 0} \frac{f(n)}{2^{n+1}} = \frac{1}{\alpha - \beta}\left(\frac{\alpha}{b - \alpha} - \frac{\beta}{b - \beta}\right) = \frac{b}{b^2 - b - 1},$$

hence the result $(*)$ for $b = 10$.

Editorial comment. Several solvers remarked that $\Sigma g(n)2^{-n}$ must be $\Sigma 2 \cdot 10^{-k}$ since the kth decimal digit of 2^n exceeds 4 if and only if the nth binary digit of $2 \cdot 10^{-k}$ is 1. B. M. M. de Weger (The Netherlands) wrote out the decimal expansion $2^n = \Sigma a_k(n)10^k$, defined $A_k(n)$ to be 1 if $a_k(n) \geqslant 5$ and 0 otherwise, and deduced Levine's result from the identity

$$\frac{2^{4 \cdot 5^{k-1}} - 1}{5^k} = \sum_{n=k}^{4 \cdot 5^{k-1} + k - 1} A_k(n) 2^{4 \cdot 5^{k-1} - n + k - 1}.$$

We also note that the reader may find it of interest to contrast this "advanced" problem with the "elementary" problem E3353 whose solution is also in this issue.

Solved also by William Armacost & Terence Shore & Douglas B. Tyler, David Callan, Curtis Cooper, Alan Frank, Joachim Herzog & Thomas Maxsein & Paul R. Smith (Germany), Ilias Kastanas, Kee-Wai Lau (Hong Kong), Eugene Levine, O. P. Lossers (The Netherlands), Reiner Martin (Germany), Thomas McDonald, Ian McGee (Canada)& Cecil Rousseau, Paul Monsky, Allen J. Schwenk, A. C. Thompson (Canada), Jack V. Wales, Jr., and B. M. M. de Weger. One incomplete solution and one incorrect solution were received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Basic Hypergeometric Series. By George Gasper and Mizan Rahman, Encyclopedia of Mathematics and Its Applications, Great Britain, 1990, xx + 287 pp.

GEORGE E. ANDREWS
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

Reviewers for the MONTHLY are told to write "a chatty essay [that]... does *not* have to be closely tied to the book under review." At the risk of violating this quite sensible suggestion, I shall begin by saying, "I love this book! It is great!"

I want to emphasize how strongly I feel about this achievement because both the book and, indeed, the entire subject can be misunderstood and dismissed by casual and brief perusal. First of all, the title, while fully consistent with historical usage, can, perhaps, mislead the outsider. The word "Basic" does not refer to fundamental or elementary. Rather it refers to the ubiquitous parameter "q" which underlies the entire subject. The next casual mistake is to flip through the pages of the book (in under five minutes) which are filled with gigantic formulas written in unfamiliar code. The natural response here is: "If this is 'basic' hypergeometric series, please, God, save me from 'advanced' hypergeometric series!"

What is being discussed in this book anyway? Recall the classical Gaussian hypergeometric series

$$1 + \sum_{n=1}^{\infty} \frac{a(a+1)\cdots(a+n-1)b(b+1)\cdots(b+n-1)}{n!c(c+1)\cdots(c+n-1)} t^n = {}_2F_1\left(\begin{matrix} a, b; \\ c \end{matrix} t\right). \tag{1}$$

These series and their generalizations are characterized by the rising factorial $a(a+1)\ldots(a+n-1)$. The q-analog or basic analog is defined by replacing the ordinary rising factorial by the basic or q-rising factorial:

$$(1-a)(1-aq)\cdots(1-aq^{n-1}) = (a;q)_n. \tag{2}$$

Thus

$${}_2\phi_1\left(\begin{matrix} a, b; \\ c \end{matrix} q, t\right) = 1 + \sum_{n=1}^{\infty} \frac{(a;q)_n (b;q)_n}{(q;q)_n (c;q)_n} t^n. \tag{3}$$

More generally

$${}_r\phi_s\left(\begin{matrix} a_1, a_2, \ldots, a_r; \\ b_1, \ldots, b_s \end{matrix} q, t\right) = \sum_{n=0}^{\infty} \frac{(a_1;q)_n (a_2;q)_n \cdots (a_r;q)_n}{(q;q)_n (b_1;q)_n \cdots (b_s;q)_n} \left[(-1)^n q^{\binom{n}{2}}\right]^{1+s-r} t^n. \tag{4}$$

Well you may ask, "What sort of interesting results can you obtain for these things?"

I'm glad you asked. For example, we find at the bottom of page 42 of the book under review

$$
{}_8\phi_7\left(\begin{matrix} a, q\sqrt{a}, -q\sqrt{a}, b, c, d, e, f; q, \dfrac{a^2q^2}{bcdef} \\ \sqrt{a}, -\sqrt{a}, \dfrac{aq}{b}, \dfrac{aq}{c}, \dfrac{aq}{d}, \dfrac{aq}{e}, \dfrac{aq}{f} \end{matrix}\right)
$$

$$
= \frac{(aq, aq/de, aq/df, aq/ef; q)_\infty}{(aq/d, aq/e, aq/f, aq/def; q)_\infty} \times {}_4\phi_3\left(\begin{matrix} aq/bc, d, e, f; q, q \\ aq/b, aq/c, def/a \end{matrix}\right)
$$

$$
+ \frac{(aq, aq/bc, d, e, f, a^2q^2/bdef, a^2q^2/cdef; q)_\infty}{(aq/b, aq/c, aq/d, aq/e, aq/f, a^2q^2/bcdef, def/aq; q)_\infty}
$$

$$
\times {}_4\phi_3\left(\begin{matrix} aq/de, aq/df, aq/ef, a^2q^2/bcdef; q, q \\ a^2q^2/bdef, a^2q^2/cdef, aq^2/def \end{matrix}\right). \tag{5}
$$

To most mathematicians such a formula is completely opaque. To the deep insider marvellous symmetries peep out here and there; however, selling this to the rest of the world as a formula whose intrinsic beauty justifies its existence is a bit much.

However, (5) does contain some special cases that have elicited some of the most rhapsodic prose ever used to describe mathematics. In particular, if f is replaced by q^{-N} where N is a positive integer and then b, c, d, e and $N \to \infty$ with $a \to 1$, we find (after invocation of Jacobi's triple product, another special case of (5))

$$
\prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})} = 1 + \sum_{n=1}^{\infty} \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}. \tag{6}
$$

Equation (6) and a companion identity were utilized by Ramanujan to deduce some amazing continued fraction formulas, e.g.

$$
\cfrac{1}{1 + \cfrac{e^{-2\pi}}{1 + \cfrac{e^{-4\pi}}{1 + \ddots}}} = \left\{\sqrt{\left(\frac{5+\sqrt{5}}{2}\right)} - \frac{1+\sqrt{5}}{2}\right\} e^{\frac{2\pi}{5}}. \tag{7}
$$

This and related formulae were contained in Ramanujan's first letter to G. H. Hardy. Hardy, in his biographical chapter on Ramanujan [7], discusses this letter at length. He begins, "I should like you to begin by trying to reconstruct the immediate reactions of an ordinary professional mathematician who receives a letter like this from an unknown Hindu clerk."

Of (7) and two related continued fraction identities he said, "[they] defeated me completely; I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class. They must be true because, if they were not true, no one would have had the imagination to invent them. Finally (you must remember that I knew nothing whatever about Ramanujan, and had to think of every possibility), the writer must be completely honest, because great mathematicians are commoner than thieves or humbugs of such incredible skill."

6. G. Gasper, q-Extensions of Clausen's formula and of the inequalities used by de Branges in his proof of the Bieberbach, Robertson, and Millin Conjectures, *S.I.A.M.* J. Math. Anal., 20 (1989) 1019–1034.
7. G. H. Hardy, Ramanujan, Cambridge University Press, Cambridge, 1940 (reprinted: Chelsea, New York, 1978).
8. S. Ramanujan, Notebooks, Vol. II, Tata Inst. of Fundamental Research, Bombay, 1957 (reprinted: Narosa, New Delhi, 1984).

Abstract and Concrete Categories. By Jiří Adámek, Horst Herrlich and George Strecker, John Wiley and Sons, New York, 1990, xii + 482 pp.

GABRIELE CASTELLINI
Department of Mathematics, University of Puerto Rico, Mayaguez, Puerto Rico 00708

Category theory is a relatively young area of mathematics that started to develop in the 1940's and early 1950's as a means of providing a general theory of mathematical structures. In doing so, it unified many existing mathematical concepts (e.g., *product*) and it made other somewhat hazy concepts (e.g., *naturality*) more precise. Its first significant uses were in the areas of algebraic topology and homological algebra, but soon afterwards its usefulness as a general language caused it to be welcomed into other classical areas of mathematics such as topology and algebra. By now it has ingressed into virtually all areas, although its major applications in analysis have been slower to develop than in many other fields. There is also a recent strong tendency to use category theoretical techniques in theoretical computer science.

But just what is category theory?

To keep this discussion on a conversational level, we will refrain from giving a technically accurate and detailed definition of a category. In order to convey some general ideas about the topic, we can say that a category consists of a class of objects and a (possibly empty) set of morphisms for each pair of objects. We usually denote by $f\colon X \to Y$ a morphism with domain the object X and with codomain the object Y. An associative operation called *composition* is defined on the class of all pairs of morphisms that satisfy the following matching condition: morphisms f and g have a *composite* (denoted $g \circ f$) provided that the codomain of the morphism f agrees with the domain of the morphism g. Each object comes with an identity morphism that forms composites with other morphisms in the only reasonable way that an identity could.

Objects of different categories can be quite different in nature. The morphisms of a given category have usually been chosen in order to interact nicely with its objects. The terminology *concrete category* is used when there is a given "underlying" category (such as the category of sets) and objects of the category in question are "based on" objects of the underlying category. Some familiar examples will illustrate this idea. Set (resp. **Top, Haus, Grp, TopGrp, Vec, Ban,** and $\mathbf{Ban}_1$) usually denotes the category whose objects are sets (resp., topological spaces, Hausdorff spaces, groups, topological groups, vector spaces, Banach spaces, and Banach spaces) and whose morphisms are functions (resp. continuous functions, group homomorphisms, continuous homomorphisms, linear transformations, linear contractions, and bounded linear maps). Each of these is an abstract category. But, each also can be considered as a concrete category based on Set, since each has objects that are sets with structure and in each case the morphisms are functions

6. G. Gasper, q-Extensions of Clausen's formula and of the inequalities used by de Branges in his proof of the Bieberbach, Robertson, and Millin Conjectures, *S.I.A.M.* J. Math. Anal., 20 (1989) 1019–1034.
7. G. H. Hardy, Ramanujan, Cambridge University Press, Cambridge, 1940 (reprinted: Chelsea, New York, 1978).
8. S. Ramanujan, Notebooks, Vol. II, Tata Inst. of Fundamental Research, Bombay, 1957 (reprinted: Narosa, New Delhi, 1984).

Abstract and Concrete Categories. By Jiří Adámek, Horst Herrlich and George Strecker, John Wiley and Sons, New York, 1990, xii + 482 pp.

GABRIELE CASTELLINI
Department of Mathematics, University of Puerto Rico, Mayaguez, Puerto Rico 00708

Category theory is a relatively young area of mathematics that started to develop in the 1940's and early 1950's as a means of providing a general theory of mathematical structures. In doing so, it unified many existing mathematical concepts (e.g., *product*) and it made other somewhat hazy concepts (e.g., *naturality*) more precise. Its first significant uses were in the areas of algebraic topology and homological algebra, but soon afterwards its usefulness as a general language caused it to be welcomed into other classical areas of mathematics such as topology and algebra. By now it has ingressed into virtually all areas, although its major applications in analysis have been slower to develop than in many other fields. There is also a recent strong tendency to use category theoretical techniques in theoretical computer science.

But just what is category theory?

To keep this discussion on a conversational level, we will refrain from giving a technically accurate and detailed definition of a category. In order to convey some general ideas about the topic, we can say that a category consists of a class of objects and a (possibly empty) set of morphisms for each pair of objects. We usually denote by $f\colon X \to Y$ a morphism with domain the object X and with codomain the object Y. An associative operation called *composition* is defined on the class of all pairs of morphisms that satisfy the following matching condition: morphisms f and g have a *composite* (denoted $g \circ f$) provided that the codomain of the morphism f agrees with the domain of the morphism g. Each object comes with an identity morphism that forms composites with other morphisms in the only reasonable way that an identity could.

Objects of different categories can be quite different in nature. The morphisms of a given category have usually been chosen in order to interact nicely with its objects. The terminology *concrete category* is used when there is a given "underlying" category (such as the category of sets) and objects of the category in question are "based on" objects of the underlying category. Some familiar examples will illustrate this idea. Set (resp. **Top, Haus, Grp, TopGrp, Vec, Ban,** and $\mathbf{Ban}_1$) usually denotes the category whose objects are sets (resp., topological spaces, Hausdorff spaces, groups, topological groups, vector spaces, Banach spaces, and Banach spaces) and whose morphisms are functions (resp. continuous functions, group homomorphisms, continuous homomorphisms, linear transformations, linear contractions, and bounded linear maps). Each of these is an abstract category. But, each also can be considered as a concrete category based on Set, since each has objects that are sets with structure and in each case the morphisms are functions

between the underlying sets that in some sense "preserve" the structure. However, notice that **TopGrp** can also naturally be viewed as a concrete category over **Top**, and as a concrete category over Grp. These three concrete categories of topological groups have quite different properties from one another, and indeed are all different from the abstract category of topological groups. Notice also that Banach spaces constitute the objects of two quite natural, but different, concrete categories based on sets.

It should also be pointed out that the definition allows categories that are far different from these examples. Objects of a category need not have elements, and morphisms need not be functions. As an example, any partially ordered set can be viewed as an abstract category whose objects are the members of the set, and $a \leqslant b$ means that there is precisely one morphism from a to b. Indeed, any such partially ordered set also can be viewed as a concrete category over the trivial category that has just one object and one morphism. This example demonstrates the strong connection between order theory and concrete category theory.

We see, then, that proofs in category theory cannot be obtained by considering points within objects. They must be based solely on the morphisms and how they are composed. However, this constraint imposed by the abstraction can often be an advantage, since it may give special insights and since any categorical result will apply in all categories. The abstraction can build quite quickly. For example, categories themselves may be viewed as objects of (quasi)categories, where the morphisms between them are called *functors*, and functors can be considered as objects, where the morphisms between them are called *natural transformations*.

The most apparent feature of category theory is the language that it provides to precisely describe mathematical concepts that otherwise would remain quite vague. This helps us to understand that, although different areas of mathematics might at first look very different from one another, they may share common features. For instance, products of topological spaces are technically different from products of groups and from products of vector spaces. If $(X_i)_{i \in I}$ is a family of topological spaces, then their topological product ΠX_i with projections $(\pi_j\colon \Pi X_i \to X_j)_{j \in I}$, satisfies the following condition: for any topological space Y and for any family of continuous functions $(f_j\colon Y \to X_j)_{j \in I}$, there exists a unique continuous function $\langle f_i \rangle\colon Y \to \Pi X_i$ such that $\pi_j \circ \langle f_i \rangle = f_j$, for every $j \in I$.

If, in the above condition, we replace "topological spaces" by "groups" (resp. "vector spaces") and replace "continuous functions" by "group-homomorphisms" (resp. "linear transformations"), we obtain the notion of direct product of groups (resp. product of vector spaces). With category theory all of these notions of products (and of course many more) can be obtained at once as the categorical *product* by taking the above condition and replacing "topological spaces" by "objects" and replacing "continuous functions" by "morphisms."

In a similar way scores of concepts in individual abstract and concrete categories have suitable categorical analogues for which they are special cases—sometimes even those concepts that seem to be defined entirely in terms of points.

For instance, we say that a topological space X is Hausdorff if any two of its distinct points have disjoint neighborhoods. This formulation clearly makes no sense outside of topology. However a characterization of Hausdorff spaces in terms of the closed diagonal is suitable for generalization to any sufficiently nice category, since it makes use only of categorical concepts such as diagonal morphisms and closure operators.

An obvious but useful consequence of this generality is that any (concrete) categorical result gives rise to a number of results, one for each particular (concrete) category and only the proof of the categorical one is needed.

The Duality Principle for categories pushes this idea even further. Any concept in a category can be dualized by basically reversing the direction of the morphisms in the category. The interesting consequence is that any true statement in category theory will yield a true dual statement. In other words, one can obtain two results (whose specializations to particular categories might seem quite unrelated) by proving just one of them.

The book *Abstract and Concrete Categories* by J. Adámek, H. Herrlich and G. E. Strecker makes these ideas especially clear. It presents the fundamental concepts of category theory by starting with the most basic ones and then gradually moving to the more advanced ones. Thus, the chosen approach is to present the special instances of abstract concepts first and the more general ones later. Such a method is bound to be more pedagogically successful than the opposite approach of giving the most general definitions first. For example, although the important, yet abstract, notion of adjoint situation comes in the latter half of the book, it comes as the culmination of three separate approaches—via reflections, via factorization structures, and via limits. Note that categorical product, mentioned above, is just one special case of limit.

This book is different from other category theory books in that approximately half of it is devoted to concrete categories, and the theories of abstract categories and concrete categories are developed roughly in parallel. Other texts tend to treat concrete categories minimally, if at all.

Besides providing the basic definitions as well as numerous examples that clarify them, the text contains many recent developments in category theory—especially within the realm of concrete categories. This makes *Abstract and Concrete Categories* suitable both for the student who is looking for a gentle introduction to the topic of category theory and for the researcher who needs a good reference work. It has been written with the same pleasant style as the previous volume by the last two authors, entitled *Category Theory*. However, it includes many important topics that were absent in the earlier text. Some of these are factorization structures for sources, injective hulls, equations and implications, cartesian closed categories, quasitopoi, topological universes, monads, and essentially algebraic, topologically algebraic, topological, and solid categories (together with corresponding functors and structure theorems for them). The reader should notice that at the time the previous book was published, in 1973, the study of many of these topics either had not yet been born, or was still in its infancy. This book, which includes many of the latest developments of the discipline previously only accessible in the research literature, was certainly needed.

To conclude, we would like to mention that the text has extensive and useful indices and that each section ends with a collection of exercises and a list of suggestions for further reading. Worth noticing is the fact that a few humorous illustrations appear now and then throughout the book to break the formality and to make the reading more light and enjoyable. We would have liked to have seen more of them.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P, L. *Liberal Learning and the Arts and Sciences Major, Volumes 1 and 2.* Association of American Colleges. *Volume 1: The Challenge of Connecting Learning,* 1990, xiii + 33 pp, [ISBN: 0-911696-51-2]; *Volume 2: Reports From the Fields,* 1991, xiii + 226 pp, (P). [ISBN: 0-911696-50-4] Reports on a study of twelve undergraduate majors focused on connections (both within and beyond the major), on context (as an essential component of liberal education), and on coherence (principles of organization that are apparent to students). The mathematics chapter is an abridged version of the report that appeared in the November 1990 issue of *Focus.* LAS

General, S. *The Mathematical Funfair.* Brian Bolt. Cambridge University Pr, 1989, vi + 122 pp, $10.95 (P). [ISBN: 0-521-37743-9] Over a hundred diverse puzzles on familiar themes, including matchstick and coin puzzles, railway shunting problems, number puzzles, chessboard puzzles, magic squares, tricks, and games. Includes detailed commentary and solutions. LCL

General, P, L. *The Popularization of Mathematics.* Eds: A.G. Howson, J.-P. Kahane. ICMI Study Ser. Cambridge University Pr, 1990, xi + 210 pp, (P). [ISBN: 0-521-40867-9; 0-521-40319-7] Reports from a September 1989 international conference held in Leeds, U.K. to study various approaches to popularizing mathematics—in the media, in schools, in museums, in books and magazines, in TV and films, in competitions. Presentations illustrate various examples and diverse objectives from Australia, Canada, China, Czechoslovakia, Denmark, France, Hungary, Italy, New Zealand, Spain, United Kingdom, and United States. Includes description of a major traveling exhibition, the "Pop Maths Roadshow," whose opening was part of the Leeds workshop. LAS

Elementary, T(13: 1). *Intermediate Algebra: Concepts and Applications, Third Edition.* Marvin L. Bittinger, Mervin L. Keedy, David Ellenbogen. Addison-Wesley, 1990, xvi + 700 pp, $29.56. [ISBN: 0-201-17296-8] Features early introduction of graphing, functions, systems of equations, and a 5-step algorithm for problem solving. This edition (a revision of *A Problem Solving Approach to Intermediate Algebra, Second Edition,* TR, January 1988) adds coverage of order of operations, algebra of functions, determinants, more inequalities, and new exercises. JNC

Precalculus, T(13: 1). *Trigonometry for College Students, Fifth Edition.* Karl J. Smith. Brooks Cole, 1991, xv + 410 pp, $42.75. [ISBN: 0-534-13728-8] This edition includes all the fundamental identities in the first chapter, new sections on inverse functions and conics in polar form, and sets of miscellaneous problems at the end of each chapter. (*First Edition,* TR, June-July 1977; *Second Edition,* TR, October 1980; *Third Edition,* TR, November 1984.) JNC

Precalculus, T(13: 1). *Fundamentals of College Algebra, Third Edition.* Charles D. Miller, Margaret L. Lial, David I. Schneider. Scott Foresman, 1990, xv + 509 pp, $32. [ISBN: 0-673-38638-4] Coverage of traditional topics includes examples and exercises designed to show how algebra techniques are used in calculus; this edition includes new historical comments and "exercises that test understanding." (*First Edition,* TR, October 1982; *Second Edition,* TR, October 1986.) JNC

Precalculus, T(13: 1), S, C. *Graphing Calculator and Computer Graphing Laboratory Manual, Second Edition.* Franklin Demana, *et al.* Addison-Wesley, 1991, iv + 118 pp, $6.36 (P). [ISBN: 0-201-55436-4] A companion manual to the Demana and Waits series of textbooks on precalculus mathematics, which emphasize calculator and computer-produced graphs. This manual teaches readers how to use several common graphing utilities, including Texas Instruments, Casio, Sharp, and Hewlett-Packard graphing calculators, and the Apple II, IBM, and Macintosh versions of *Master Grapher.* All told, a valuable compendium of information, and for students, a useful alternative to full owner's manuals. PZ

Precalculus, T, P. *Approaching Precalculus Mathematics Discretely: Explorations in a Computer En-*

vironment. Philip G. Lewis. MIT Pr, 1990, xiv + 557 pp, $24.95 (P). [ISBN: 0-262-12138-7] Develops the concepts of functions, linear transformations, vectors, recursion, limits and trigonometric functions using Logo; an appendix introduces Logo. JNC

Finite Mathematics, T(13-14), S, C. *Problem Solver for Finite Mathematics and Calculus.* Kenneth L. Wiggins. PWS-Kent, 1991, xiv + 221 pp, $16 (P). [ISBN: 0-534-92439-5] *Problem Solver* is an IBM-PC package (supplied on two 5.25 inch disks; not tested by this reviewer) of mathematical routines: graphics (2-d plots of functions and tables), matrix manipulations, statistical computations, numerical integration, and root-finding, etc. The book itself is both a user manual and a compendium of problems (on precalculus, finite mathematics, calculus, elementary linear algebra, and statistics) to which the software can be applied. Each section includes at least one example worked in detail. PZ

Education, L. *Sourcebook for Science, Mathematics, and Technology Education, 1990-1991.* Eds: Betty J. Calinger, Barbara Walthall. AAAS, 1990, v + 202 pp, $12.95 (P). Information on organizations, offices, agencies, and people who are working to improve science and mathematics education. Includes state and national organizations, public and private groups, professional associations and education agencies, books and journals. Indexed by name, organization, state, and publication. Coverage of mathematics is slim: no mention of the NCTM *Standards* nor of NRC's *Everybody Counts.* A useful reference for names and addresses. LAS

Education, P, L*. *Mathematics and Cognition.* Eds: Pearla Nesher, Jeremy Kilpatrick. ICMI Study Ser. Cambridge University Pr, 1990, 180 pp, $39.50. [ISBN: 0-521-36608-9] A concise, well-referenced state-of-the-art survey of research in the psychology of mathematics education (PME) prepared over a two-year period by the International Group for PME which is affiliated with the International Commission for Mathematics Instruction (ICMI). Five chapters that advance by age level, from arithmetic to calculus, are enclosed by survey chapters on epistemology and psychology and on future perspectives for research. LAS

Education, P*, L. *They're Not Dumb, They're Different: Stalking the Second Tier.* Sheila Tobias. Research Corp (6840 E. Broadway Blvd., Tucson, AZ 85710-2815), 1990, 94 pp, (P). Eyewitness accounts of typical introductory university physics and chemistry courses prepared by experienced observers—humanities graduate students who volunteered to take the courses and keep detailed notes of their experiences, observations, and feelings. Examples: "The lack of community ...produces a totally passive classroom experience." "My class is full of intellectual warriors ...taught to fear cooperation." "Students key into buzz-words and catch phrases ...They are not turned on by ideas." "Everyone, including the professor, is geared toward exams." "There are lots of atomized individuals ...physics is not a place where people befriend one another." Mathematicians can learn much about how "second tier" students—those who enter college not already committed to science—react to typical introductory courses. LAS

Education, P. *Black Students and the Mathematics, Science, and Technology Pipeline: Turning the Trickle into a Flood.* Howard University Pr. *The Journal of Negro Education,* 59:3 (Summer 1990), 281 pp, (P). A special "Yearbook Issue" devoted to science and mathematics education, particularly about mathematics and especially about algebra—which for most minority students is the barrier to the mathematics and science pipeline. Includes research studies (e.g., on "learned helplessness," on the effects of Eurocentrism on motivation of Afro-American and Hispanic students), and project reports (including one by Jaime Escalante). A rich resource of ideas, experience, and people. LAS

History, P, L*. *Regiomontanus: His Life and Work.* Ernst Zinner. Transl: Ezra Brown. Stud. in History & Philos. of Math., V. 1. North-Holland (US Distr: Elsevier Science), 1990, x + 402 pp, $82. [ISBN: 0-444-88792-X] An extraordinarily detailed account of the scholarly life of this unheralded precursor to Copernicus, also known as Johannes Molitoris (or Müller) of Königsberg. Discoverer of the fifth and sixth perfect numbers, Regiomontanus (1436-1476) was the pre-eminent astronomer of his age, tabulating and correcting astronomical tables and working with kings and popes on calendar corrections. Regiomontanus' planetary observations laid the foundation for *De revolutionibus* without ever challenging church geocentric doctrine. LAS

Foundations, P, L.** *Brouwer's Intuitionism.* Walter P. van Stigt. Stud. in History & Philos. of Math., V. 2. North-Holland (US Distr: Elsevier Science), 1990, xxvi + 530 pp, $94.75. [ISBN: 0-444-88384-3] A fascinating intellectual biography tracing the roots of Brouwer's philosophical views from his angry, arrogant youth through a calculated detour into topology (to establish mathematical *bona fides*), and back to foundational studies. Includes carefully documented expositions of Brouwer's philosophical views, his analysis of language and logic, and his intuitionist mathematics. Appendices provide original text and translations of several important unpublished papers and correspondence. Reveals clearly how personalities combined with intellectual passions to fuel the "foundation wars" in the first part of the twentieth century. LAS

Graph Theory, S, P. *Fuzzy Sets in Information Retrieval and Cluster Analysis.* Sadaaki Miyamoto. Theory & Decision Library, Ser. D, V. 4. Kluwer Academic, 1990, x + 259 pp, $99. [ISBN: 0-7923-0721-6] Introduction to fuzzy set theory for finite sets. Connects these ideas to theory of finite, fuzzy graphs. Terse but thorough exposition of major results. Accessible to readers with no background in fuzzy set theory. Remainder of the text divides into

independent chapters applying fuzzy set theory to information retrieval (in particular, bibliographic retrieval) and hierarchical cluster analysis. Includes many simple examples and a large section devoted to various algorithms of clustering schemes. MK

Graph Theory, T(16-17: 2), P. *Decompositions of Graphs.* Juraj Bosák. Math. & Its Applic., V. 47. Kluwer Academic, 1990, xviii + 248 pp, $112. [ISBN: 0-7923-0747-X] Begins with basic concepts and definitions and proceeds rapidly to an in-depth study of decompositions of graphs into isomorphic subgraphs. While prerequisites are quite minimal, a certain degree of mathematical sophistication is necessary. Extensively referenced, it includes a 36-page bibliography of relevant research papers. Conjectures and open problems sprinkled throughout. AD

Discrete Mathematics, T(13-14: 1, 2). *Foundations of Discrete Mathematics.* Peter Fletcher, Hughes Hoyle, C. Wayne Patty. PWS-Kent, 1991, xviii + 781 pp, $35. [ISBN: 0-534-92373-9] Begins with elementary logic, methods of proof, sets, functions, induction, and recursion. Other chapters examine algorithms, counting, Boolean algebra, equivalence relations, order, and formal languages. A chapter on matrices constitutes a brief course in linear algebra. Treatment of graph theory and trees is extensive (171 pages). Advanced chapter on combinatorics discusses generating functions. No calculus background necessary, only two years of high school algebra. AD

Discrete Mathematics, T(13: 1). *Discrete Mathematics for Computing.* A. Vince, C. Morris. Math. & Its Applic. Ellis Horwood (US Distr: Simon & Schuster), 1990, 297 pp, (P). [ISBN: 0-13-217522-3] Assumes little formal mathematical background. Covers sets, logic, data representation and manipulation, graph theory and associated algorithms, relations, functions, combinatorics, probability, and error correcting codes, all at a very elementary level. Examples of algorithms presented in a language-free manner. Answers for exercises supplied. AD

Number Theory, S(17-18), P, L. *Representation Theory and Automorphic Functions.* I.M. Gel'fand, M.I. Graev, I.I. Pyatetskii-Shapiro. Transl: K.A. Hirsch. Academic Pr, 1990, xviii + 426 pp, $39.50 (P). [ISBN: 0-12-279506-7] Reprint of a 1968 translation. A valuable resource for anyone venturing into the Langlands program. BC

Linear Algebra, T(16-17). *Extreme Properties of Linear Transformations.* Ali R. Amir-Moéz. Polygonal, 1990, vii + 166 pp, $24. [ISBN: 0-936428-12-0] Hermitian, positive, unitary properties of transformations on inner-product spaces. Exercises; terse style. MLR

Group Theory, P. *Combinatorial Group Theory.* Eds: Benjamin Fine, Anthony Gaglione, Francis C.Y. Tang. Contemp. Math., V. 109. AMS, 1990, xii + 191 pp, $37 (P). [ISBN: 9-8218-5116-0] The proceedings from the AMS special session on combinatorial group theory and infinite groups held at the University of Maryland, April 1988. The papers in this volume represent a mix of pure research papers and expository papers; together they convey a sense of the direction in the field. LCL

Algebra, S(18), P. *Automata and Algebras in Categories.* JiříAdámek, Věra Trnková. Math. & Its Applic. Kluwer Academic, 1990, xiv + 474 pp, $167. [ISBN: 0-7923-0010-6] Chapters 1 and 2 introduce the basic concepts of sequential and tree automata leading into work of Arbib and Manes. Succeeding chapters treat free algebras, minimal realization and reduction, universal realization, and nondeterministic automata. Chapter notes, references, index. JS

Algebra, P. *Computers in Algebra.* Ed: Martin C. Tangora. Lect. Notes in Pure & Appl. Math., V. 111. Marcel Dekker, 1988, viii + 162 pp, $89.75 (P). [ISBN: 0-8247-7975-4] Papers from a conference on computers in algebra held at the University of Illinois at Chicago in December 1985. Most of the papers are on finite groups and such problems as enumerating cosets, finding efficient presentations, constructing explicit representations, and computing cohomology. JAS

Algebra, T(16-17: 1), L. *Steps in Commutative Algebra.* R.Y. Sharp. London Math. Soc. Stud. Texts, V. 19. Cambridge University Pr, 1990, xi + 321 pp, $49.50; $17.95 (P). [ISBN: 0-521-39338-8; 0-521-39732-4] A gentle introduction, assuming only basic abstract algebra. Some exercises. JD-B

Algebra, P. *Lecture Notes in Mathematics-1428: Blocks of Tame Representation Type and Related Algebras.* Karin Erdmann. Springer-Verlag, 1990, xv + 312 pp, $32.20 (P). [ISBN: 0-387-52709-5] Studies algebras associated to blocks of tame representation type. These are the 2-blocks whose defect groups are dihedral or semi-dihedral or (generalized) quaternion. LC

Calculus, T(13-14: 2, 3).** *Calculus.* Gilbert Strang. Wellesley-Cambridge Pr (Box 82-279, Wellesley, MA 02181), 1991, xv + 655 pp, $50 (Libraries). [ISBN: 0-9614088-2-0] A lively fresh look at calculus by a master expositor: conventional topics cast in new light and used for important purposes. Employs several standard examples (e.g., speedometer/odometer, electrocardiogram) throughout to provide continuous grounding in real problems. Introduces computer and calculator methods as well as mathematical ideas rooted in these tools (e.g., iteration, chaos from Newton's method, patterns in the graph of $\sin n$ for integral $n, 0 \leq n \leq 10000$). Contains, as expected, several thousand exercises, including regular "fill-in-the-blank" passages in lieu of chapter summaries. Emphasizes *pairs* of functions $f(t)$ and $v(t)$ (distance and velocity), differential equations (beginning in Chapter 2), and visualization (graphs and zooms). The real innovation, however, is in the exposition: direct, refreshing, enthusiastic, personal. LAS

Real Analysis, P. *Analysis II: Convex Analysis and Approximation Theory.* Ed: R.V. Gamkrelidze. Encyclop. of Math. Sci., V. 14. Springer-Verlag,

1990, 255 pp, $59.50. [ISBN: 0-387-18179-2] Basic ideas of convex analysis (subspaces, convex cones, duality, linear and convex programming, etc.) and approximation theory (classical problems and methods, and "best" approximation by subspace, Fourier, linear and nonlinear methods). Extensive reference section. MLR

Complex Analysis, P. *Lecture Notes in Mathematics-1435: Computational Methods and Function Theory.* Eds: St. Ruscheweyh, *et al.* Springer-Verlag, 1990, vi + 211 pp, $22 (P). [ISBN: 0-387-52768-0] Seventeen papers on aspects of complex functions theory—conformal mapping, univalent functions, extremal problems—comprise partial proceedings of a March 1989 conference held at Valparaiso, Chile. Computer and numerical methods appear frequently throughout the papers. PZ

Partial Differential Equations, T(17-18), P. *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis, Second Edition.* Lars Hörmander. Grund. der math. Wissenschaften, B. 256. Springer-Verlag, 1990, xi + 440 pp, $69, [ISBN: 0-387-52343-X]; Study Edition, $39.50 (P). This edition includes exercises with answers and hints. (*First Edition,* TR, March 1984; Extended Review, December 1985.) MLR

Partial Differential Equations, P. *Singular Perturbations I: Spaces and Singular Perturbations on Manifolds without Boundary.* Leonid S. Frank. Stud. in Math. & Its Applic., V. 23. North-Holland (US Distr: Elsevier Science), 1990, xxiv + 555 pp, $89.75. [ISBN: 0-444-88134-4] The first of two volumes devoted to singular perturbations, treating here linear perturbations on smooth manifolds without boundary. Chapter 1 develops the necessary background in functional analysis and distributions; Chapter 2 discusses Sobolev spaces; and Chapter 3 deals with elliptic and hyperbolic perturbations, using a variety of techniques. Notes, extensive bibliography. JS

Partial Differential Equations, P. *Nonlinear Evolution Equations That Change Type.* Eds: Barbara Lee Keyfitz, Michael Shearer. Inst. for Math. & Its Applic., V. 27. Springer-Verlag, 1990, xiv + 284 pp, $35. [ISBN: 0-387-97353-2] Many simple models of fluid flow, including granular flow, multiphase flow, and flow in porous media, involve partial differential equations that change type. The papers in this volume, based on the 1988-89 IMA Program on Nonlinear Waves, address numerical and analytical difficulties posed by change of type. JO

Partial Differential Equations, C. *PLTMG: A Software Package for Solving Elliptic Partial Differential Equations, Users' Guide 6.0.* Randolph E. Bank. Frontiers in Appl. Math., V. 7. SIAM, 1990, xiv + 164 pp, $22.50 (P). [ISBN: 0-89871-255-6] A reference manual for a software package for solving a general class of scalar, parameter dependent, elliptic partial differential equations in general regions of the plane. JNC

Functional Analysis, T(18: 1, 2), S, P. *Functional Analysis: An Introduction for Physicists.* Nino Boccara. Academic Pr, 1990, xiii + 327 pp, $44.50. [ISBN: 0-12-108810-3] Burdened by neither "paralyzing rigor" or "barren formalism" the author's intent is to make the basic concepts of modern functional analysis intelligible and useful for theoretical physicists. Many (but not all) proofs are given, with problem sets (and solutions) to enhance understanding. Chapters treat measure and intregration, Lebesgue spaces, Hilbert spaces, distributions, and linear operators. References, index. JS

Functional Analysis, T(17-18), S, P, L. *C^*-Algebras and Operator Theory.* Gerard J. Murphy. Academic Pr, 1990, x + 286 pp, $44.50. [ISBN: 0-12-511360-9] Contains an accessible introduction to core material (spectral theory, C^*-algebras, Hilbert space operators, ideals and positive functions, von Neumann algebras, representations of C^*-algebras), plus the treatment of topics of high contemporary profile (including C^*-tensor products and K-theory). The careful presentation, choice of topics, and number of exercises make this useful as a course text. KS

Functional Analysis, P. *Analysis of Toeplitz Operators.* Albrecht Böttcher, Bernd Silbermann. Springer-Verlag, 1990, 512 pp, $79. [ISBN: 0-387-52147-X] A Toeplitz operator $T(a) : H^p \to H^p$ for $a \epsilon L^\infty$ and $1 < p < \infty$ is defined by $f \to P(af)$ where P is the Riesz projection. These operators have uses in (for example) physics, probability, and information theory as well as in operator theory and function theory. The book is intended for both beginners and specialists. MLR

Analysis, P. *Mathematical Structures of Nonlinear Science: An Introduction.* Melvyn S. Berger. Nonlinear Topics in Math. Sci., V. 1. Kluwer Academic, 1990, xii + 421 pp, $99. [ISBN: 0-7923-0728-3] A book that grew out of four lectures the author gave at the University of California: New ideas on integrability of nonlinear systems; vortex phenomena in fluids (a classic topic that stands to benefit from new mathematical ideas of nonlinear science); new connections between global differential geometry and nonlinear analysis; superconductivity—the simplest nonlinear gauge theory that works. AWR

Analysis, T(18), S, P, L. *The Analytical and Topological Theory of Semigroups: Trends and Developments.* Eds: Karl Heinrich Hofmann, Jimmie D. Lawson, John S. Pym. Walter de Gruyter, 1990, xi + 398 pp, DM 138. [ISBN: 0-89925-698-8] A comprehensive survey of the current state of research in the theory of semigroups and its applications. This volume emphasizes the relationships with topology and analysis (functional and harmonic), and is presented in four parts: Lie theory and algebraic geometry; structure of compact semigroups; functional analysis on semigroups; applications to systems theory, number theory, probability, and topology. KS

Algebraic Geometry, P. *Sheaves on Manifolds.* Masaki Kashiwara, Pierre Schapira. Grund. der math. Wissenschaften, B. 292. Springer-Verlag, 1990, x + 512 pp, $89. [ISBN: 0-387-51861-4] Intended to present a self-contained exposition of sheaf

theory in the context of derived categories. Includes a short history (in French) of the subject. JO

Differential Geometry, T(18), S, P. *An Introduction to CR Structures.* Howard Jacobowitz. Math. Surveys & Mono., No. 32. AMS, 1990, x + 237 pp, $74. [ISBN: 0-8218-1533-4] Very readable introduction to the subject of CR structures. Begins with counterexamples to the Riemann mapping theorem in higher dimensions involving CR manifolds. Focuses on the equivalence problem as solved by Cartan and by Chern and Moser. Concludes by showing the nonsolvability of the Lewy operator. Accessible to anyone who has taken a graduate course in differential geometry. OJ

Differential Geometry, T(18: 1), S, P. *Submanifolds and Isometric Immersions.* Marcos Dajczer, *et al.* Math. Lect. Ser., V. 13. Publish or Perish, 1990, viii + 173 pp, $26. [ISBN: 0-914098-22-5] Text for advanced graduate course in differential geometry. Uses flat bilinear forms to prove rigidity theorems. Assumes knowledge of graduate differential geometry, but reviews many of the basic facts and theorems. Well-written and well-referenced. Exercises, bibliography. OJ

Algebraic Topology, P. *Classical Aspherical Manifolds.* F. Thomas Farrell, L. Edwin Jones. CBMS Reg. Conf. Ser. in Math., No. 75. AMS, 1990, vii + 54 pp, $24 (P). [ISBN: 0-8218-0726-9] Lectures from the NSF-CBMS Conference on K-theory and Dynamics held in Gainesville, Florida in January 1989. Includes an epilogue with results obtained since the conference. SP

Topology, P. *Open Problems in Topology.* Eds: Jan van Mill, George M. Reed. North-Holland (US Distr: Elsevier Science), 1990, xiv + 692 pp, $92.25. [ISBN: 0-444-88768-7] Designed as a source book for thesis problems and a challenge to top researchers. Offers a collection of unsolved conjectures and problems in various areas of topology (e.g., general and point-set, computer science oriented, geometric and algebraic, etc.) submitted by researchers in each of the fields. Description of each problem and of related results and conjectures is given. Indexes for easy reference. An excellent resource. OJ

Dynamical Systems, P. *Decision Processes in Dynamic Probabilistic Systems.* Adrian V. Gheorghe. Math. & Its Applic., V. 42. Kluwer Academic, 1990, xvii + 354 pp, $112. [ISBN: 0-7923-0544-2] Brings together theory of Markov and semi-Markov processes, dynamical and linear programming, and decision process theory with either complete or partial information. Mathematics is advanced, examples are scarce, notation is non-standard, and exposition is terse. Includes Fortran programs to calculate transition probabilities, first passage times, and other characteristics of Markov processes. MK

Dynamical Systems, P. *Nonlinear Structures in Physical Systems: Pattern Formation, Chaos, and Waves.* Eds: Lui Lam, Hedley C. Morris. Springer-Verlag, 1990, x + 331 pp, $49. [ISBN: 0-387-97344-3] Collection of thirty-five short papers from the Second Woodward Conference held in November 1989 at San Jose State University. Topics include pattern formation of bubbles; viscous fingers; chemical systems; frustrated systems; chaos; self-organized criticality; traveling waves, and various nonlinear structures in magnetic systems, plasmas, polymers, thin films, and biological systems. SP

Dynamical Systems, T(17), S*, P, L. *Perspectives of Nonlinear Dynamics, Volume 2.* E. Atlee Jackson. Cambridge University Pr, 1990, xviii + 632 pp, $59.50. [ISBN: 0-521-35458-7] Extensive introduction written by a physicist for graduate students and researchers in the sciences and engineering. Many helpful illustrations, historical notes, exercises, and examples. Topics in this volume include maps in R^2 and T^2, twist maps, Lorenz, KAM, Lyapunov exponents, solitons, KdV, and much more. SP

Dynamical Systems, P. *Mathematics of Nonlinear Science.* Ed: Melvin S. Berger. Contemp. Math., V. 108. AMS, 1990, xi + 153 pp, $32 (P). [ISBN: 0-8218-5114-4] Proceedings of AMS Special Session held in January 1989 at Phoenix, Arizona. Papers deal primarily with new ideas about non-chaotic phenomena in physical science. Topics include instability, global extensions of KAM, integrable systems, bifurcation analysis in fluids, and gauge theory. SP

Dynamical Systems, P. *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems.* B. Hanzon. Mathematisch Centrum, 1989, [ISBN: 90-6196-371-0]. *Part I,* CWI Tract, V. 63, 223 pp, Dfl. 58 (P); *Part II,* CWI Tract, V. 64, 413 pp, Dfl. 48 (P). Results of the author's research over the past ten years on the problem of model identification. More specifically, among all possible models that might describe a certain phenomenon, how does one choose the most nearly correct one? In Part I the author considers representations of linear dynamical systems, deterministic and stochastic, and describes these as abstract objects. He imposes a topology so that one can decide when models are close and when far apart. The set of stochastic linear systems of a fixed degree form a differentiable manifold under certain conditions. Part II investigates an algorithm that uses overlapping parametrizations to investigate this model set. LCL

Control Theory, P. *Nonlinear Controllability and Optimal Control.* Ed: H.J. Sussmann. Pure & Appl. Math., V. 133. Marcel Dekker, 1990, viii + 472 pp, $115. [ISBN: 0-8247-8258-5] Contributions to a 1987 workshop in Finite Dimensional Controllability and Optimal Control held at Rutgers University, this volume is intended to serve as an overview of recent efforts to put new tools of differential geometric control theory to work on such old control theory problems as optimal trajectories, system invertability, sampling, and more. AWR

Probability, P. *Lecture Notes in Mathematics-1426: Séminaire de Probabilités XXIV, 1988/89.* Eds: J. Azéma, P.A. Meyer, M. Yor. Springer-Verlag, 1990, v + 490 pp, $47.90 (P). [ISBN: 0-387-52694-3] Conference proceedings; contains 39 pa-

pers, 15 in English, the rest in French. RSK

Probability, T(18), P. *Entropy and Information Theory.* Robert M. Gray. Springer-Verlag, 1990, xxiii + 332 pp, $53.50. [ISBN: 0-387-97371-0] A well-written, research-level treatment of the theory of probability information measures and their application to coding theorems for information sources and noisy channels. Begins with basic definitions as originally introduced by Shannon. Covers standard stationary ergodic theorems and rates, divergence, relative entropy, and ergodic theorems for densities. Many of the examples and applications are to Markovian systems. A large section of examples of channels and codes, as well as chapters on distortion, source coding theorems, and coding for noisy channels. Large reference list. MK

Probability, P. *Continuous Bivariate Distributions, Emphasising Applications.* T.P. Hutchinson, C.D. Lai. Rumsby Scientific (POB 76, Rundle Mall, Adelaide, South Australia 5000), 1990, xxxi + 412 pp, $63. [ISBN: 0-7316-7206-2] Comprehensive compilation of probability distributions involving two variables, including ways they are generated. Extensive collection of examples of their use in modelling data. RSK

Stochastic Processes, T(16-17), S, L. *Probability, Statistics, and Queueing Theory With Computer Science Applications, Second Edition.* Arnold O. Allen. Comput. Sci. & Scientific Comput. Academic Pr, 1990, xxi + 740 pp, $49.95. [ISBN: 0-12-051051-0] Truly three books in one: probability theory, statistical inference, and queueing theory as they apply to computer science problems. Sections are terse, but cover standard topics, include good examples, and contain a wide variety of exercises. Queueing theory section covers models for computer systems such as Jackson networks and BCMP Queueing Networks. Makes use of several software packages: MINITAB, SAS/STAT, EXPLORE, APL, and Mathematica. An appendix contains Mathematica programs to assist understanding of queueing models. Philosophy is clearly how theorems and theory are used in practical computer science problems. (*First Edition,* TR, March 1979.) MK

Computational Statistics, T(16-17), L. *An Introduction to Generalized Linear Models.* Annette J. Dobson. Chapman & Hall, 1990, x + 174 pp, $52.50. [ISBN: 0-412-31100-3] An undergraduate-graduate level treatment of generalized linear models. Very accessible exposition full of examples. Incorporates both MINITAB and GLIM examples. Standard topics: exponential families, generalized linear models, estimation, inference (mostly likelihood-based), multiple regression (includes diagnostics, model selection, non-linearity), ANOVA, ANCOVA, logistic regression, and log-linear models. MK

Computational Statistics, S. *How SAS Works: A Comprehensive Introduction to the SAS System, Second Edition.* Paul A. Herzberg. Springer-Verlag, 1990, xvii + 193 pp, $29 (P). [ISBN: 0-387-97291-9] Written both as an introduction and as a bridge to the more extensive reference manuals published by the SAS Institute. Concentrates on the DATA steps used to input, create, and modify data, while limiting the discussion of PROC (procedure) steps to descriptive statistics. RSK

Statistics, P*. *Asymptotics in Statistics: Some Basic Concepts.* Lucien Le Cam, Grace Lo Yang. Ser. in Stat. Springer-Verlag, 1990, viii + 180 pp, $29. [ISBN: 0-387-97372-9] Updated rewriting of a set of lectures given by Le Cam in 1968, originally published in French. Concentrates on local asymptotic normality (LAN) conditions, but now also includes a discussion of the locally asymptotically mixed normal case (LAMN). Each of the six non-introductory chapters concludes with interesting historical remarks. RSK

Statistics, T(16-17: 1, 2), P. *Statistical Design and Analysis of Experiments with Applications to Engineering and Science.* Robert L. Mason, Richard F. Gunst, James L. Hess. Ser. in Prob. & Math. Stat. Wiley, 1989, xvi + 692 pp, $64.95. [ISBN: 0-471-85364-X] Nontheoretical practitioner's guide, assuming only college algebra. Stresses the strategy of experimentation, data analysis, and the interpretation of experimental results. Main sections are on experimental design (presented separately from the analysis), analysis of data from designed experiments, and regression modeling. Includes such topics as fractional factorial experiments, response-surface designs, multiple comparison procedures, analysis of covariance, and outlier detection. RSK

Statistics, P. *Proceedings of the Thirty-Fifth Conference on the Design of Experiments.* US Army Research Office (POB 12211, Research Triangle Park, NC), 1989, xii + 327 pp, (P). Proceedings from the thirty-fifth conference on the Design of Experiments in Army Research, Development and Testing, which resiliently survived the San Francisco-area earthquake in October 1989. Includes a special session on statistical issues related to combat modelling. Other invited topics include statistical graphics, an overview of statistical methods for categorical data, and an appraisal of several multistage selection procedures. MK

Statistics, T*(17: 1), C, L. *Regression Analysis: Theory, Methods, and Applications.* Ashish Sen, Muni Srivastava. Texts in Stat. Springer-Verlag, 1990, xv + 347 pp, $39. [ISBN: 0-387-97211-0] Up-to-date account of the theory and methods of regression analysis, assuming a background of mathematical statistics. First part covers least squares regression assuming the Gauss-Markov conditions. Second part presents methods to alleviate the effects of violations of these conditions. Third part discusses multicollinearity, variable search, and biased estimation. Includes a computer disk which contains most of the over 50 data sets used in the text to illustrate the material. RSK

Statistics, P. *Statistical Methods in Longitudinal Research.* Ed: Alexander von Eye. Stat. Modeling & Decision Sci. Academic Pr, 1990, $74.95

each. *Volume I: Principles and Structuring Change,* xiv + 273 pp, [ISBN: 0-12-724960-5]; *Volume II: Time Series and Categorical Longitudinal Data,* xiv + 328 pp. [ISBN: 0-12-724962-1] Collection of contributed papers from methodological researchers of longitudinal analyses in the social sciences. Numerous examples and applications including discussions of computer implementation. Topics: change measurement, missing data estimation, alternative analyses of repeated-measure designs, longitudinal factor analysis, latent variable structural equation models, scaling, analysing event histories, linear and nonlinear curve fitting, multivariate time series models, descriptive and associative developmental models, and analysis of categorical longitudinal data. MK

Statistics, P*. *U-Statistics: Theory and Practice.* A.J. Lee. Stat.: Textbooks & Monographs, V. 110. Marcel Dekker, 1990, xi + 302 pp, $79.75. [ISBN: 0-8247-8253-4] "Reasonably comprehensive" description of the theory of U-statistics (which include many of the common statistics in use today) that has accumulated over the past forty years, with particular attention to asymptotic theory. Also includes a selection of applications of this theory. RSK

Computer Literacy, S. *Calculator Logic Systems and Mathematical Understandings.* Enid R. Burrows. NCTM, 1990, viii + 133 pp, $22.50 (P). [ISBN: 0-87353-295-3] Detailed discussion with numerous examples of the major calculator operating systems—algebraic logic (keys executed in order pressed), operator logic (operations preserve logical hierarchy), and reverse Polish notation (operators act as functions on input stack). Focuses solely on button-pushing with little discussion of the important implications of these diverse systems for mathematics or pedagogy. LAS

Algorithms, T(17-18: 2), S, P. *Combinatorial Algorithms for Integrated Circuit Layout.* Thomas Lengauer. Ser. in Comput. Sci. Wiley, 1990, xxviii + 697 pp, $59.95. [ISBN: 0-471-92838-0] Introduction to, the theory of, and then latest results on combinatorial problems on integrated circuit layout and their algorithmic solutions. Unified theory and results for both CAD practitioners and algorithm theoreticians. Includes graph algorithms, optimization techniques, operations research, and statistics results as needed for layout optimization; also includes actual layout problems and solutions like global and detailed routing. JPH

Computer Systems, T(15-17: 1), L*.** *Operating System Concepts, Third Edition.* Abraham Silberschatz, James L. Peterson, Peter B. Galvin. Addison-Wesley, 1991, xviii + 696 pp, $47.50. [ISBN: 0-201-51379-X] A substantially revised and updated edition of a popular textbook that focuses on fundamental concepts. In addition to bringing some of the older material up-to-date and adding new exercises, the authors have added new chapters on distributed file systems and on the Mach operating system. AO

Artificial Intelligence, S(18), P. *Artificial Intelligence in Process Engineering.* Ed: Michael L. Mavrovouniotis. Academic Pr, 1990, xiv + 367 pp, $49.95. [ISBN: 0-12-480575-2] A collection of papers illustrating the application of artificial intelligence methods to chemical and biochemical process engineering. The goal of the editor is to acquaint process engineers and artificial intelligence practitioners with each others' tools and problems. The desired result is more active collaboration between these two communities. SM

Artificial Intelligence, P, L. *Connectionist Robot Motion Planning: A Neurally-Inspired Approach to Visually-Guided Reaching.* Bartlett W. Mel. Perspectives in A.I., V. 7. Academic Pr, 1990, xvi + 165 pp, $29.95. [ISBN: 0-12-490020-8] This book describes MURPHY, a robot arm and camera system whose task is to reach for objects while avoiding obstacles. The software controlling MURPHY's motion is based on a connectionist model of learning by experience. JO

Applications, S(15-16). *Life Insurance Mathematics.* Hans U. Gerber. Transl: W. Neuhaus. Springer-Verlag, 1990, xiii + 131 pp, $49.50. [ISBN: 0-387-52944-6] Translation of a 1986 Swiss monograph. Concise introduction to the mathematics of life contingencies, assuming some familiarity with probability theory. RSK

Applications (Biological Science), S(18), L. *Mathematical Ecology of Plant Species Competition: A Class of Deterministic Models for Binary Mixtures of Plant Genotypes.* Anthony G. Pakes, Ross A. Maller. Stud. in Math. Biology. Cambridge University Pr, 1990, xiii + 193 pp, $49.50. [ISBN: 0-521-37388-3] Presents two mathematical (not computer) models describing the interaction of two species of annual plants. The existence of both "soft" and "hard" seeds plays an integral role in these models. SM

Applications (Electrical Engineering), T(18: 2), P, L. *Semiconductor Equations.* P.A. Markowich, C.A. Ringhofer, C. Schmeiser. Springer-Verlag, 1990, x + 248 pp, $54.50. [ISBN: 0-387-82157-0] A text on the drift diffusion equations modelling the charge transport in semi-conductors. The book arrives at these equations via a unified discussion of the hierarchy of semi-conductor models starting with the kinetic transport equations. The material is addressed to applied mathematics, electrical engineers, and solid state physicists. Includes problems and extensive references. MPR

Applications (Engineering), P. *Proceedings of the Third European Conference on Mathematics in Industry.* Eds: John Manley, Sean McKee, David Owens. European Consort. for Math. in Industry, V. 5. BG Teubner Stuttgart (US Distr: Kluwer Academic), 1990, xii + 564 pp, $127. [ISBN: 0-7923-0807-7] Motivated by a desire to acquaint more mathematicians with the needs and opportunities for people adept at using mathematical models in industry, the European Consortium for Mathematics in Industry was formed. This book reflects the inclusiveness of the organization: vibrations of overhead

transmission lines, temperature surges in thermistors, spiral turns by means of forming cutters, optimization in large gas networks, dairy management, and more. There are several general articles on such topics as supercomputers, simulation, etc. AWR

Applications (Engineering), P. *Mathematics in Industrial Problems, Part 3.* Avner Friedman. Math. & Its Applic., V. 31. Springer-Verlag, 1990, xiii + 187 pp, $24. [ISBN: 0-387-97436-9] The third in a series produced by the Institute of Mathematics and Its Applications at the University of Minnesota. The seventeen problems in this volume have been collected by the author with two criteria in mind. The problems listed have a bona fide application to a manufacturing problem, and simultaneously present an opportunity for mathematical novelty. Each description is short, giving a definition of the problem, a brief overview of current techniques with references, and a list of relevant open questions. SM

Applications (Engineering), T(16-17: 1). *Variational and Finite Element Methods: A Symbolic Computation Approach.* A.I. Beltzer. Springer-Verlag, 1990, xi + 254 pp, $59.50. [ISBN: 0-387-51598-4] An introduction to variational and finite element methods for undergraduate engineering students. It is distinguished by its use of a dialect of MACSYMA as a tool that allows the student to focus on the ideas presented. AO

Applications (Physics), T(17-18: 1), S, P, L. *The Quantum Statistics of Dynamic Processes.* E. Fick, G. Sauermann. Transl: W.D. Brewer. Ser. in Solid-State Sci., V. 86. Springer-Verlag, 1990, xiii + 395 pp, $79. [ISBN: 0-387-50824-4] Graduate-level introduction to methods and applications of statistical physics within a quantum-theoretical framework. Covers basic concepts and response and relaxation theory for small perturbations from equilibrium. Includes exercises, with solutions. BC

Applications (Physics), P. *Lecture Notes in Physics-359: Problems Involving Change of Type.* Ed: K. Kirchgässner. Springer-Verlag, 1990, xii + 207 pp, $32.40. [ISBN: 0-387-52595-5] Sixteen papers on variational problems, phase transitions and dynamics, viscoelasticity, hyperbolic-elliptic problems, and degenerate parabolic equations—in short, things that spontaneously change type. BC

Applications (Physics), S(16-18), P, L. *A Development of the Equations of Electromagnetism in Material Continua.* Harry F. Tiersten. Tracts in Natural Philos., V. 36. Springer-Verlag, 1990, xi + 156 pp, $49. [ISBN: 0-387-97241-2] Based on lecture notes for a graduate course in mechanics. Great attention is given to developing the equations of electrostatics, magnetostatics, and electromagnetism from simply-stated first principles. The interplay between the careful mathematics and the concrete boundary conditions is very illuminating. A fascinating excursion for anyone familiar with Maxwell's equations. MU

Applications (Physics), T(18: 1, 2). *Probability and Heat: Fundamentals of Thermostatistics.* Friedrich Schlögl. Friedr. Vieweg & Sohn (US Distr: GLP International Vieweg), 1989, xii + 249 pp, DM 98. [ISBN: 3-528-06343-2] Derives basic thermodynamic theory from a probabilistic analysis of molecular microdynamics, and then again from phenomenological postulates, in an attempt to resolve some of the usual conceptual confusion of thermodynamics. JO

Applications (Physics), T(17-18: 1, 2), S, P, L. *Relativistic Quantum Mechanics: Wave Equations.* Walter Greiner. Theoretical Phys., V. 3. Springer-Verlag, 1990, xvi + 345 pp, $39.50 (P). [ISBN: 0-387-50986-0] This fluent translation of the 1987 German original begins with the Klein-Gordon equation (for spin-0 particles) and then works through the Dirac equation (for spin-1/2). Although most attention is given to Dirac's theory, other allied topics such as spinors and Lorentz invariance are also covered. The final chapters are devoted to the wave equations for particles with higher spin. MU

Applications (Simulation), T(18), P. *Computational Vision.* Harry Wechsler. Comput. Sci. & Scient. Computing. Academic Pr, 1990, xvii + 558 pp, $64.50. [ISBN: 0-12-741245-X] A lucid, interdisciplinary cognitive science approach to vision, with extensive material on parallel and distributed processing and neural networks. Uses techniques from mathematics, physics, and neuropsychology of perception. Reviews many real applications. Extensive bibliography. For readers in computer science, cognitive science, robotics, applied mathematics. DFA

Applications (Simulation), P, L. *Developmental Systems: At the Crossroads of System Theorem, Computer Science, and Genetic Engineering.* S. Węgrzyn, J.-C. Gille, P. Vidal. Springer-Verlag, 1990, vii + 122 pp, $39.50. [ISBN: 0-387-97340-0] Developmental systems are models of genetic codes containing rules of transformation and "operating systems" grouped into a "generating word" that is capable of life-like behavior including regeneration after trauma, sensitivity to mutation, cloning, and reproduction. Offers the opportunity to simulate life in computers with a mathematical model. Captures the essence of DNA without its enormous complexity. LAS

Reviewers

DFA: David F. Appleyard, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; AD: Amy Davidow, Macalester; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; MK: Michael Kahn, St. Olaf; RSK: Richard S. Kleber, St. Olaf; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; JS: John Schue, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

Mathematical Writing



USING WRITING TO TEACH MATHEMATICS

Andrew Sterrett, Editor

...the day will come, I believe, when the value of writing to learn will be universally acknowledged

Rueben Hersh in
Writing to Learn Mathematics and Science

Need help in getting started as an individual or as a member of a department facing a Writing Across the Curriculum requirement? Learn how others have made use of student assistants, both undergraduate and graduate, in ways that benefit students and faculty members alike. Read that feedback from student journals provide early-warning signals for instructors, as well as help students clarify their own thought processes.

This collection of essays is an outgrowth of the widespread interest shown in sessions of contributed papers on writing given at the 1988 and 1989 Annual Meetings of the MAA. Many of the 30 authors of the essays included in this volume participated in those sessions and each has considerable experience in requiring students to write about mathematics.

Included in this volume are essays that:

- remind professors how frequently mathematicians, regardless of their careers, are asked to write, ("Mathematicians Write; Mathematics Students Should Too"),
- provide a theoretical framework by which to assess writing assignments ("Writing for Educational Objectives in a Calculus Course"), and
- give practical examples of assignments that work ("Writing in Mathematics: A Plethora of Possibilities").

This source book is filled with practical suggestions. It will enhance the comprehension that your students have of mathematics.

160 pp., Paperback, 1990,
ISBN-0-88385-066-4

List $12.50

Catalog Number NTE-16

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

When was the last time a computer program helped you think about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Undercover Applications Four Times A Year



We collect the latest mathematical applications and put them under one cover four times a year. **The UMAP Journal** blends lesson-length, undergraduate-level teaching modules, professional commentary, and contemporary articles for practitioner and professor alike.

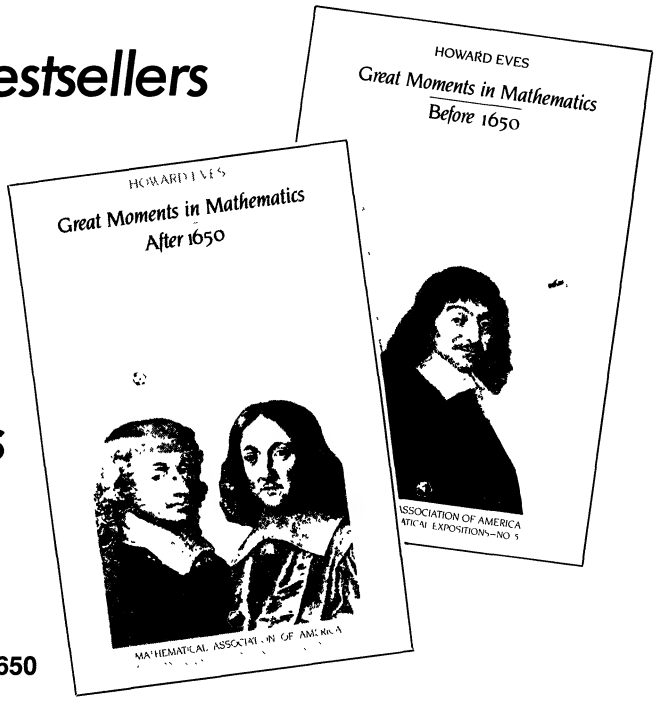
The UMAP Journal
Tackling The Problems Of A New Age

Individual subscription (annual)	United States	Foreign
Journal only	$37/$34*	$49/$46*
Journal and *UMAP Modules: Tools for Teaching*	$51/$48*	$64/$61*

* Membership in COMAP or any of the cooperating societies entitles the subscriber to a $3 discount.

COMAP, Inc. 60 Lowell Street, Arlington, MA 02174

The MAA Bestsellers on the History of Mathematics



GREAT MOMENTS IN MATHEMATICS BEFORE 1650

Howard Eves

This outstanding volume (along with its companion Volume 7) has become an MAA bestseller. It contains 23 fascinating mathematical episodes from antiquity to about 1650.

270 pp., Paper, 1982, ISBN 0-88385-310-8

List: $17.00 MAA Member: $13.00

Catalog Number DEP-05

GREAT MOMENTS IN MATHEMATICS AFTER 1650

Howard Eves

This is a companion volume to Volume 5.

270 pp., Paper, 1982, ISBN 0-88385-311-6

Paper List: $17.00 MAA Member: $13.00
Catalog Number DEP-07 (Paper)

Hardbound List : $27.00
MAA Member: $20.00
Catalog Number DOL-07 (Hardbound)

Both books available in paperbound editions as a package: List: $31.50 MAA Member: $23.00

Catalog Number DEP-057

Howard Eves made a valuable contribution to the Dolciani Mathematical Exposition series . . . The twenty lectures included are a delight to read. They place each "great moment" in its historical context and lay special emphasis on human aspects of each achievement. No algebra or geometry teacher should be without this book.

Tom Walsh in *The Mathematics Teacher*

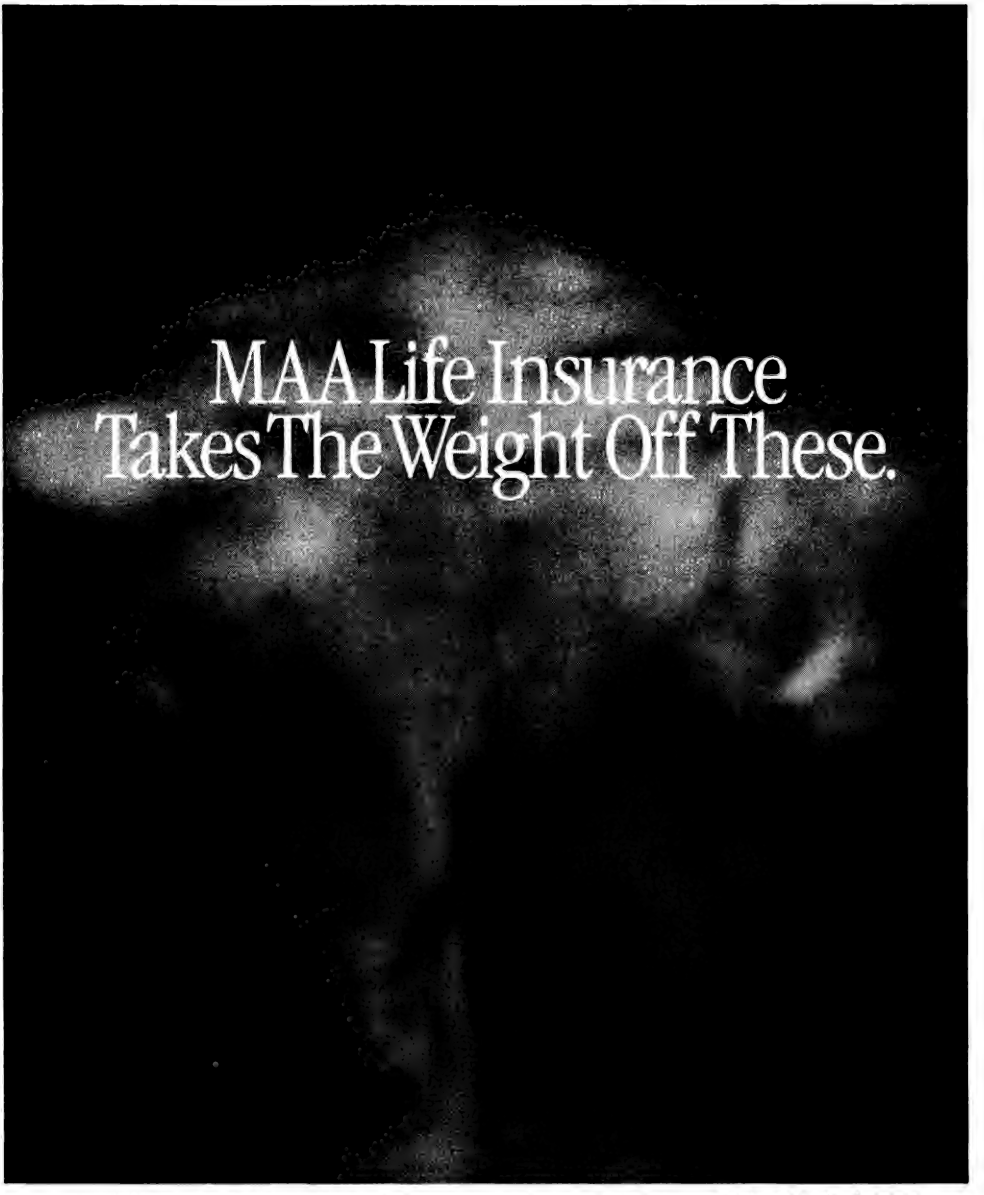
He has the knack of bringing ancient mathematics alive, at the same time relating it to modern ideas and developments. His work is readable and informative and indeed bears reading a second time. His exposition is clear and he is not content with generalities, but gives actual examples of early methods.

E.H. Lockwood in
The Mathematical Gazette

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036



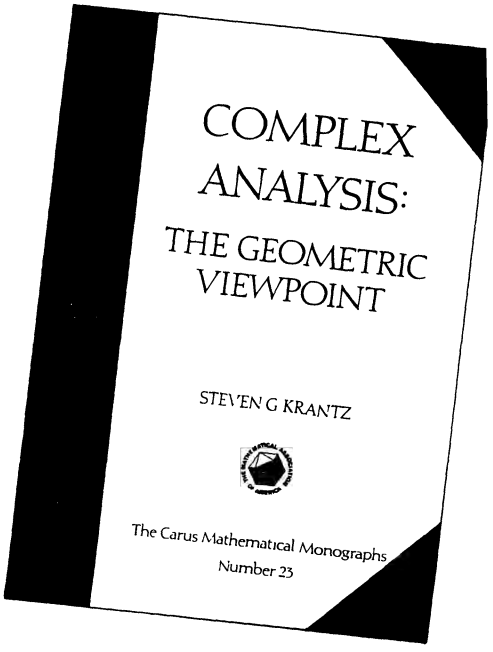
Your Association uses group purchasing power to negotiate top quality Term Life Insurance, at a very low price. What's more, it's insurance specially designed for your professional needs. It can be tailored for the individual, and it can stay with you even if you change jobs.

To take advantage of this benefit of membership—and take the weight off your shoulders—call 1-800-424-9883 for further details (in Washington, D.C. call 457-6820).

MAA INSURANCE
Designed by Members.
For Members.

The MAA Life Plan is underwritten by Connecticut General Life Insurance Company, A CIGNA Company, Hartford, Connecticut 06152 on form number GM 3000

New From Steven G. Krantz



COMPLEX ANALYSIS:
THE GEOMETRIC VIEWPOINT

Steven G. Krantz

Geometric methods have been used in complex analysis since the 1930s when Lars Ahlfors discovered that they give a nice way to look at the Schwarz lemma. Since that time they have become a central part of the research activities of complex analysis. However these important techniques have never found their way into a text accessible to a broad audience.

Steven G. Krantz, a leading worker in complex analysis and a well-known mathematical expositor, has written the first book explaining how complex analysis can be studied using methods of geometry. Assuming no background in Riemannian geometry, and only one semester of complex analysis, Krantz explains the role of Hermitian metrics and of curvature in understanding the Schwarz lemma, normal families, Picard's theorems, conformal mappings, and many other topics. A minimum of geometric formalism is used to gain a maximum of geometric and analytic insight. The climax of the book is an introduction to several complex variables from the geometric viewpoint. Poincaré's theorem, that the ball and bidisc are biholomorphically inequivalent, is discussed and proved.

Except for the minimal background requirements, the book is self-contained. A review of relevant topics in the classical theory of one complex variable is provided. The style is light and inviting. The book is a must for anyone with an interest in complex analysis. Take a glance at the main chapter headings and order your copy today

- Principal Ideas of Classical Function Theory,
- Basic Notions of Differential Geometry,
- Curvature and Applications,
- Some New Invariant Metrics,
- A Glimpse of Several Complex Variables

210 pp., 1990, Hardbound,
ISBN 0-88385-026-5

List $22.00 MAA Member $18.50

Catalog Number CAM-23

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

HISTORY AND THEORY IN MATHEMATICS

Games and Economic Behavior

Editor
Ehud Kalai
Northwestern University, Evanston, Illinois

Games and Economic Behavior publishes original and survey papers dealing with game-theoretic modeling in the social, biological, and mathematical sciences. Papers published are mathematically rigorous as well as accessible to readers in related fields. The purpose of the journal is to facilitate cross-fertilization between the theory and application of game-theoretic reasoning.

Volume 3 (1991), 4 issues **ISSN 0899-8256**
In the U.S.A. and Canada: $110.00 **All other countries: $133.00**

Historia Mathematica

Editor
Eberhard Knobloch
Technische Universität Berlin, Germany
Managing Editor
David E. Rowe
Pace University, Pleasantville, New York

Historia Mathematica is concerned with the history of all aspects of the mathematical sciences in all parts of the world and all historical periods. The journal publishes occasional biographies of mathematicians and historians, studies of organizations and institutions, essays on historiography, and articles on the interactions among all facets of mathematical activity and other aspects of culture and society.

Published under the Auspices of the International Commission on the History of Mathematics of the Division of the History of Science of the International Union of the History and Philosophy of Science

Volume 18 (1991), 4 issues **ISSN 0315-0860**
In the U.S.A. and Canada: $85.00 **All other countries: $105.00**

Under New Editorial Direction

Journal of Approximation Theory

Editors-in-Chief
Paul Nevai
The Ohio State University, Columbus
Allan Pinkus
Technion, Haifa, Israel

The ***Journal of Approximation Theory*** is devoted to new advances in pure and applied approximation theory and related areas. The journal covers classical and abstract approximation, constructive approximation, harmonic analysis and Fourier expansions, function spaces and interpolation of operators, real and complex function theory, orthogonal polynomials and special functions, general orthogonal systems, approximation theoretical aspects of difference and differential equations, multivariate approximation, and spline functions of one and several variables.

Volumes 64–67 (1991), 12 issues **ISSN 0021-9045**
In the U.S.A. and Canada: $452.00 **All other countries: $534.00**

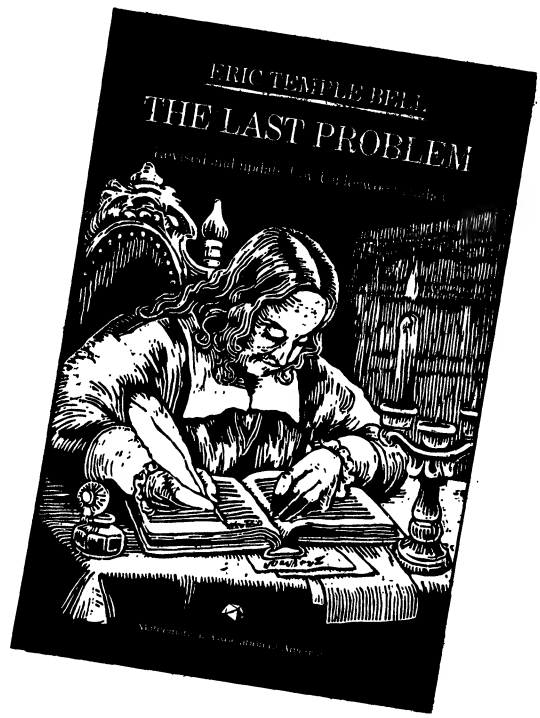
Sample copies and privileged personal rates are available upon request.
For more information, please write or call:

ACADEMIC PRESS, INC.
Journal Promotion Department
1250 Sixth Avenue
San Diego, CA 92101, U.S.A.
(619) 699-6742

All prices are in U S dollars and are subject to change without notice

S1151

Revised and Updated



THE LAST PROBLEM

E. T. Bell
Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990
ISBN-0-88385-451-1

List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

Just Published from HBJ!

INTRODUCTORY ALGEBRA
and
INTERMEDIATE ALGEBRA
RONALD HATTON and
GENE R. SELLERS
both of *Sacramento City College*
Each Paperbound, 1991

ELEMENTARY ALGEBRA
Second Edition
and
INTERMEDIATE ALGEBRA
Second Edition
YOSHIKO YAMATO and
MARY JANE CORDÓN
both of *Pasadena City College*
Each Hardcover, 1991

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN,
Drexel University, and
ARNOLD SHAPIRO
Hardcover, 1991

FINITE MATHEMATICS
WILLIAM OWEN and
FRED CUTLIP
both of *Central Washington University*
Hardcover, 1991

When you asked for a sensible, no-frills, mainstream calculus textbook at a reduced cost for your students, we heard you...

CALCULUS
One and Several Variables
ROBERT ELLIS and
DENNY GULICK
both of *University of Maryland*
Hardcover, 1991 $29.95 (Net)!

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of *Vassar College*
Hardcover, 1991

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.,
Michigan State University
Hardcover, 1991

ABSTRACT ALGEBRA
DENNIS KLETZING,
Stetson University
Hardcover, 1991

HBJ

HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

For College Mathematics Teachers

A SOURCE BOOK FOR COLLEGE MATHEMATICS TEACHING

Alan Schoenfeld, Editor.
Prepared by the Committee on the Undergraduate Teaching of Mathematics

Do you want a broader, deeper, more successful mathematics program? This Source Book points to the resources and perspectives you need.

This book provides the means for improving instruction, and describes the broad spectrum of mathematical skills and perspectives our student should develop. The curriculum recommendations section shows where to look for reports and course resources that will help you in your teaching. Extensive descriptions of advising programs that work is included, along with suggestions for teaching that describe a wide range of instructional techniques. You will learn about how to use computers in your teaching, and how to evaluate your performance as well as that of your students.

Every faculty member concerned about teaching should read this book. Every administrator with responsibility for the quality of mathematics programs should have a copy.

80 pp., 1990, Paper,
ISBN 0-88385-068-0

List $10.00

Catalog Number SRCE

ORDER FROM



The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

This book will make a difference

CALCULUS *by Gilbert Strang*

A direct and friendly textbook, written for students to read. The mathematics is alive, the explanations are clear, the length is reduced. This book fully supports calculus courses at all institutions.

So much time and effort go into our teaching of calculus – now we have a forward-looking book. You will see fresh ideas and a new spirit. The strength of calculus comes through *actively.* Your students will enjoy working with this textbook.

Previous texts by Gilbert Strang:
Linear Algebra and Its Applications (HBJ, 3rd edition)
Introduction to Applied Mathematics (Wellesley-Cambridge Press)

Please ask to see Strang's Calculus

Text: 16 chapters, 615 pages
January 1991

Wellesley-Cambridge Press
Box 82-279
Wellesley MA 02181
→ (617)431-8488

NEW! VERSION 2

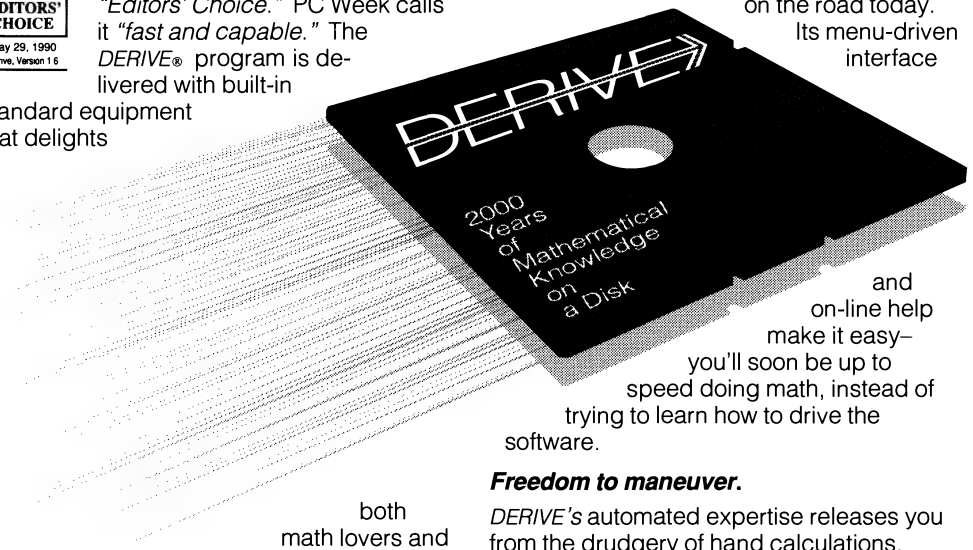
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE May 29, 1990 Derive, Version 1.6

DERIVE® – A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

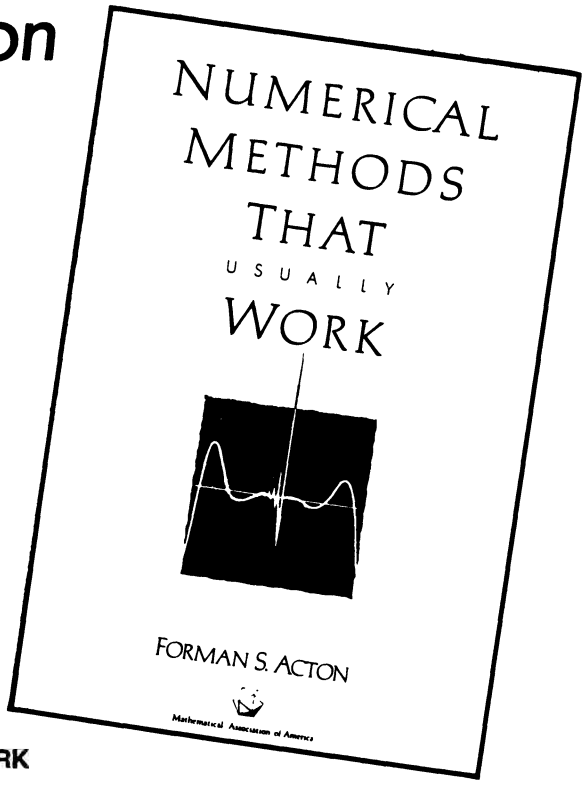
And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

New From Forman S. Acton



NUMERICAL METHODS THAT WORK

Forman Acton

Numerical Methods That Work, originally published in 1970, has been reissued by the MAA with a new preface and some additional problems. Acton deals with a common sense approach to numerical algorithms for the solution of equations: algebraic, transcendental, and differential. It assumes that a computer is available for performing the bulk of the arithmetic. The book is divided into two parts, either of which could form the basis of a one-semester course in numerical methods. Part I discusses most of the standard techniques; roots of transcendental equations, roots of polynomials, eigenvalues of symmetric matrices, and so on. Part II cuts across the basic tools, stressing such common problems as instabilities in extrapolation, removal of singularities, and loss of significant figures. The book is written with clarity and precision, intended for practical rather than theoretical use.

An eminently readable and very well movitivated introductory text.
Mathematics of Computation

A first-rate book which can be used either as a text or reference.
Choice

560 pp., Paperbound, 1990
ISBN 0-88385-450-3

List $16.50 MAA Member $13.00

Catalog Number NMTW

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 4 April 1991

Contents

(ISSN 0002-9890)

ARTICLES

The Search for a Finite Projective Plane of Order 10 C. W. H. LAM 305

The Fiftieth William Lowell Putnam Mathematical Competition
LEONARD F. KLOSINSKI, GERALD L. ALEXANDERSON, AND LOREN C. LARSON 319

NOTES

Symmetry Classes: Functions of Three Variables N. METROPOLIS AND GIAN-CARLO ROTA 328

Irrationality Without Number Theory RICHARD BEIGEL 332

Another Elementary Approach to the Jordan Form J. I. HALL 336

Borromean Circles Are Impossible
BERNT LINDSTRÖM AND HANS-OLOV ZETTERSTRÖM 340

The Inclusion $L^p(\mu) \subseteq L^q(\nu)$ A. G. MIAMEE 342

THE TEACHING OF MATHEMATICS

π and the Limit of $(\sin \alpha)/\alpha$ LEONARD GILLMAN 346

The Structure of Orthogonal Transformations LEE RUDOLPH 349

A Simple Proof of Zorn's Lemma JONATHAN LEWIN 353

The Converse of Liouville's Theorem RICHARD GRANT 354

Visualizing the p-adic Integers ALBERT A. CUOCO 355

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 365

Advanced Problems and Solutions 372

REVIEWS

Japanese Temple Geometry Problems. By Hidetosi Fukagawa and Dan Pedoe DAN SOKOLOWSKY 381

Symmetries of Culture: Theory and Practice of Plane Pattern Analysis
By Dorothy K. Washburn and Donald W. Crowe. DORIS SCHATTSCHNEIDER 383

TELEGRAPHIC REVIEWS 387

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $33 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

The Search for a Finite Projective Plane of Order 10

C. W. H. LAM,* *Computer Science Department, Concordia University, Montréal, Québec, Canada H3G 1M8*

CLEMENT LAM received a Ph.D. in mathematics at Caltech under Herbert Ryser in 1974. After graduation, he spent a year at the University of Calgary. In 1975, he joined the computer science department at Concordia University, where he is currently a professor. His research interest is in combinatorics, especially in the application of computers to solve mathematical problems.



Dedicated to the memory of Herbert J. Ryser

1. Prologue. When I was a graduate student looking for a thesis topic, Herbert Ryser advised me not to work on the projective plane of order 10. Even though he was extremely interested in this subject, he believed that it was too difficult and that I might get nowhere with it. I took his advice and chose another problem. Somehow, this problem has a beauty that fascinates me as well as many other mathematicians. Finally in 1980, I succumbed to the temptation and started working on it with some of my colleagues. We eventually managed to get somewhere, but unfortunately, Dr. Ryser is no longer with us to hear of the final result. This is an expository article describing the evolution of the problem and how computers were used to solve it.

2. The History of the problem. *A finite projective plane of order n*, with $n > 0$, is a collection of $n^2 + n + 1$ lines and $n^2 + n + 1$ points such that

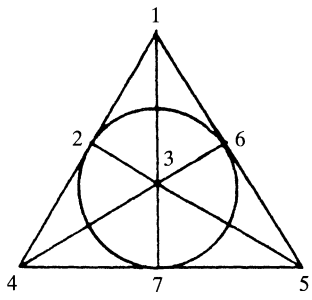
1. every line contains $n + 1$ points,
2. every point is on $n + 1$ lines,
3. any two distinct lines intersect at exactly one point, and
4. any two distinct points lie on exactly one line.

There are several different definitions for a finite projective plane and this set of axioms is chosen to highlight the striking duality of lines and points. One can interchange the words "line" and "point" in the definition and obtain essentially the same axioms! The attractiveness of these objects is in their simplicity and their reliance on the language of geometry. One is tempted to start drawing lines on paper and may soon discover some simple examples.

The smallest example of a finite projective plane is a triangle, the plane of order one. The smallest non-trivial example is of order 2, as shown in FIG. 1. There are seven points labelled from 1 to 7. There are also seven lines labelled $L1$ to $L7$. Six of them are straight lines but $L6$ is represented by the circle through the points 2, 6, and 7. The reader is invited to verify that the axioms of a finite projective plane are satisfied.

An early reference to a finite projective plane is in the paper by Veblen [**32**], which studied the axioms for geometry and used the plane of order 2 as an

*This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants A9373, A9413, 0011 and by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche under Grants EQ2369 and EQ3886.



$L1 = \{1, 2, 4\}$
$L2 = \{2, 3, 5\}$
$L3 = \{3, 4, 6\}$
$L4 = \{4, 5, 7\}$
$L5 = \{5, 6, 1\}$
$L6 = \{6, 7, 2\}$
$L7 = \{7, 1, 3\}$

FIG. 1. The finite projective plane of order two.

example. Veblen also proved that this plane of order 2 cannot be drawn using only straight lines. In a series of papers [**32**, **33**, **34**], Veblen, Bussey and Wedderburn established the existence of most of the planes of small orders, as well as all four non-isomorphic planes of order 9. One of the missing orders is $n = 6$.

In 1938, Bose [**4**] explained why there is no projective plane of order 6. He related the existence of a finite projective plane of order n to the existence of a hyper-Graeco-Latin square. The term "Graeco-Latin square" originated with Euler in 1782. In our modern terminology, they are called orthogonal Latin squares. Let us define them.

A *Latin square* of order n is an $n \times n$ matrix satisfying the following properties:

1. all the entries are integers between 1 and n,
2. in every row, no entry is repeated, and
3. in every column, no entry is repeated.

Let $S_1 = [s_{ij}^{(1)}]$ and $S_2 = [s_{ij}^{(2)}]$ denote two Latin squares of order n. They are said to be *orthogonal* provided that the n^2 2-samples $(s_{ij}^{(1)}, s_{ij}^{(2)})$ for $i, j = 1, 2, \ldots, n$ are distinct. A simple way to visualize the definition is to put the second square, slightly shifted, on top of the first square. The resulting n by n array of 2-samples has no repeated entries, and is often referred to as a *Graeco-Latin square* or an *Euler square*.

FIGURE 2 contains two examples of Latin squares of order 4. They are orthogonal and the resulting Graeco-Latin square is shown in FIG. 3. Is there a third Latin square orthogonal to both the squares in FIG. 2? Yes, there is one, as shown in FIG. 4. Can one find a fourth? The answer is no. One can easily prove the following theorem [**26**, p. 80].

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

FIG. 2. Two orthogonal latin squares of order four.

11	22	33	44
23	14	41	32
34	43	12	21
42	31	24	13

FIG. 3. A Graeco-Latin square of order 4.

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

FIG. 4. A Latin square orthogonal to those in FIG. 2.

THEOREM 1. *Let $S_1, S_2, \ldots, S_t$ be a set of t mutually orthogonal Latin squares of order $n \geq 3$. Then*

$$t \leq n - 1. \tag{1}$$

If equality holds in (1), then the orthogonal set is said to be *complete*. Finally, we are ready to state Bose's result [**26**, p. 92].

THEOREM 2. *Let $n \geq 3$. We may construct a projective plane of order n if and only if we may construct a complete set of $n - 1$ mutually orthogonal Latin Squares of order n.*

Why does Bose's result explain the non-existence of a projective plane of order 6? It states that such a plane exists if and only if there exists a complete set of 5 mutually orthogonal Latin squares of order 6. The possible existence of even a pair of orthogonal Latin squares of order 6 was a much older problem.

In a 1782 paper [**12**], Euler started by stating the problem of the 36 officers. This problem asks for an arrangement of 36 officers of 6 ranks and from 6 regiments in a square formation of size 6 by 6. Each vertical and each horizontal line of this formation is to contain one and only one officer of each rank and one and only one officer from each regiment. Euler denoted the 6 regiments by the Latin letters a, b, c, d, e, f and the 6 ranks by the Greek letters α, β, γ, δ, ε, ζ. He further remarked that the characteristic of an officer was determined by the two letters, one Latin and the other Greek, and that the problem consists of arranging the 36 combinations of two letters in a square in such manner that every row and every column contains the six Latin as well as the six Greek letters. This was the origin of the term "Graeco-Latin square." Euler observed that the first step was to arrange the Latin letters in a square so that no letter was missing either from any row or from any column. He called this square a Latin square. If he had chosen to arrange the Greek letters instead, then we would probably have Graeco squares rather than Latin squares. In any case, if we label both the ranks and the regiments from 1 through 6, then Euler's problem reduces to the construction of a pair of orthogonal Latin squares of order 6.

Euler found no solution to this particular problem. He then conjectured that no solution exists if the order of the square is of the form $n \equiv 2 \pmod 4$. This is the famous Euler's conjecture. The first case $n = 2$ is trivially impossible. Tarry around 1900 [**27**] verified by a systematic enumeration that Euler's conjecture holds for $n = 6$. Since there does not exist even a pair of orthogonal Latin squares, Bose's result implies the non-existence of a projective plane of order 6.

Yet, there is something unpleasant about a systematic hand enumeration: it is messy and it is error prone. Mathematicians did find a better explanation in the celebrated Bruck-Ryser theorem [**7**], which was published in 1949.

THEOREM 3. (Bruck-Ryser) *If* $n \equiv 1, 2 \pmod 4$, *then a necessary condition for the existence of a finite projective plane of order n is that integers* x, y *exist satisfying* $n = x^2 + y^2$.

Since $6 \equiv 2 \pmod 4$ and it is not a sum of two integer squares, the Bruck-Ryser theorem implies the non-existence of a projective plane of order 6. How did they prove the theorem? We will not repeat it here, but one of the crucial steps involved the use of an incidence matrix.

The *incidence matrix* $A = [a_{ij}]$ of a projective plane of order n is an $n^2 + n + 1$ by $n^2 + n + 1$ matrix where the columns represent the points and the rows represent the lines. The entry a_{ij} is 1 if point j is on line i; otherwise, it is 0. For example, FIG. 5 gives the incidence matrix for the projective plane of order 2. In

	1	2	3	4	5	6	7
$L1$	1	1	0	1	0	0	0
$L2$	0	1	1	0	1	0	0
$L3$	0	0	1	1	0	1	0
$L4$	0	0	0	1	1	0	1
$L5$	1	0	0	0	1	1	0
$L6$	0	1	0	0	0	1	1
$L7$	1	0	1	0	0	0	1

FIG. 5. An incidence matrix for the plane of order two.

terms of an incidence matrix, the properties of being a projective plane are translated into:

1. A has constant row sum $n + 1$,
2. A has constant column sum $n + 1$,
3. the inner product of any two distinct rows of A is 1, and
4. the inner product of any two distinct columns of A is 1.

The conditions on row sums and row inner products can be encapsuled in the following matrix equation:

$$AA^T = nI + J, \tag{2}$$

where A^T denotes the transpose of the matrix A, I denotes the identity matrix, and J the matrix of all 1's. Every diagonal entry on the right-hand side of Eq. (2) is

$n + 1$, which implies that the inner product of any row of A with itself is $n + 1$ or, equivalently, its row sum is $n + 1$. Every off-diagonal entry is 1, which is the same as requiring the inner product of any two distinct rows of A to be 1. Ryser also proved that A is a normal matrix [**25**]; in other words,

$$AA^T = A^TA.$$

Hence, Eq. (2) also implies the conditions regarding column sums and column inner products. The Bruck-Ryser theorem starts from this equation and proves that it implies n is a sum of two integer squares when $n \equiv 1, 2 \pmod 4$.

It is surprising that the Bruck-Ryser theorem has a partial converse [**13**].

THEOREM 4. *If* $n \equiv 0, 3 \pmod 4$ *or if* $n \equiv 1, 2 \pmod 4$ *and* $n = x^2 + y^2$, *then there exists a rational matrix* A *satisfying equation* (2).

Of course, if the matrix A is actually the incidence matrix of a projective plane, then it must have entries which are either 0 or 1. Although this theorem guarantees only a matrix with rational entries, it suggests that the necessary part of the Bruck-Ryser theorem is very close to being also sufficient.

Projective planes are special cases of a class of combinatorial objects called symmetric block designs. We are not going to discuss block designs, except to mention that Chowla and Ryser have generalized the Bruck-Ryser theorem to symmetric block designs [**10**], which it is now known as the Bruck-Ryser-Chowla theorem. Here again, a partial converse exists, providing more credence to the hope that the conditions in the Bruck-Ryser-Chowla theorem are both necessary and sufficient. This hope is now shattered by the non-existence of the finite projective plane of order 10.

Let us return to the history of projective planes. Now that we have a good explanation of the nonexistence of a plane of order 6, what is the next unknown case? It is $n = 10$. Since $10 = 1^2 + 3^2$, a plane of order 10 would exist if the necessary condition of the Bruck-Ryser theorem is also sufficient. On the other hand, $10 \equiv 2 \pmod 4$, and so, if one believes Euler's conjecture, then it does not exist.

First, Euler's conjecture was shown to be false. In 1959, Bose and Shrikhande [**5**] constructed a pair of orthogonal Latin squares of order 22. Then Parker [**23, 24**] constructed a pair of orthogonal Latin squares of order 10. Together they showed that Euler's conjecture is false for all orders greater than six [**6**]. This raised the hope for the existence of a plane of order 10.

History indicated that significant advances were made when one branch of mathematics was shown to be related to a different branch of mathematics. It is not surprising that the beginning of the end of the plane of order 10 occurred when people started studying the binary error-correcting code associated with it.

3. The beginning of the end. Let A be the incidence matrix of a finite projective plane of order 10. Let V be the vector space generated by the rows of A over $\mathbf{F}_2$, the finite field with two elements $\{0, 1\}$. A vector in V is called a *codeword*. The *weight* of a codeword is the number of 1's in the codeword. Let w_i be the number of codewords of weight i. We define the *weight enumerator* of V to be

$$\sum_{i=0}^{111} w_i x^i.$$

In 1970, Assmus gave a talk at Oberwolfach entitled "The Projective Plane of Order Ten?" which discussed the properties of V. After the talk, there was a lot of anticipation about this new approach. Maybe one could derive a contradiction such as showing that one of the weights is non-integral or negative. Unfortunately, no such simple contradictions were found.

However, there were several important advances. Assmus and Mattson showed [2] that the weight enumerator is uniquely determined by w_{12}, w_{15}, and w_{16}. Since their report is not readily available in most libraries, we shall refer to the paper of MacWilliams, Sloane, and Thompson [22], which proved many of the same results. One of the many innocuous but extremely useful results was:

THEOREM 5. *Let l be any line of the plane, and v any codeword of V. Then*

$$|v \cap l| \equiv |v| \pmod 2.$$

For example, every line must intersect a codeword of even weight in an even number of points. This gives us an extra condition beyond what is available from the definition of a projective plane. Roughly speaking, this condition reduces the number of possibilities for each line by half. The cumulative effect of this condition is tremendous and it is the main reason that makes an exhaustive search possible.

Furthermore, MacWilliams et al. showed that $w_{15} = 0$ after using about 3 hours of computer time on a General Electric 635. Bruen and Fisher later showed in [8], that $w_{15} = 0$ also followed from an earlier computer result by Denniston [11]. However, the method of MacWilliams, et al., illustrated how to continue attacking the problem. This can be summarized as follows:

> Given any weight i, we assume that a codeword of weight i exists. By considering the intersection patterns of a few selected lines with the i points of this codeword, we arrive at a small number of starting configurations, each corresponding to a submatrix of the incidence matrix. Then, we try to complete the rest of the incidence matrix. If we succeed, then it is time to celebrate because we have constructed a plane. If none of the starting configurations can be so completed, then the plane of order 10 does not contain any codeword of weight i and $w_i = 0$.

This method requires first the generation of all the possible starting configurations. A good reference is the 1980 paper [14] by Marshall Hall, Jr., which analyzed in detail the starting configurations for codewords of small weight $i \leq 20$. Given a starting configuration, the attempt to complete it is basically a backtracking process. The term "backtrack" was coined by D. H. Lehmer in the 1950's, but backtrack techniques have been used to solve puzzles for a long time. It is a tedious and lengthy task, one that is best suited for a computer.

3.1 **Backtrack search using a computer.** The first description of a generalized backtrack search algorithm was in [35]. Let us first define a restricted version of the search problem.

> Given a collection of sets of candidates $C_1, C_2, C_3, \ldots, C_m$ and a boolean *compatibility* function $P(x, y)$ defined for all $x \in C_i$ and $y \in C_j$, find an *m-tuple* $(x_1, \ldots, x_m)$ with $x_i \in C_i$ such that $P(x_i, x_j)$ is true for all $i \neq j$.

The m-tuple satisfying the above condition is called a *solution*. The term *compatibility* was first introduced by Carter in 1974 [**9**]. This version is restricted because the compatibility function is only defined on ordered pairs rather than on all k-tuples, $1 \leq k \leq m$.

For example, if we take $m = n^2 + n + 1$ and let C_i be the set of all candidates for column i of the incidence matrix of a projective plane, then $P(x, y)$ can be defined as

$$P(x, y) = \begin{cases} \text{true} & \text{if } \langle x, y \rangle = 1 \\ \text{false} & \text{otherwise,} \end{cases}$$

where $\langle x, y \rangle$ denotes the inner product of columns x and y. A solution is then a complete incidence matrix.

In a backtrack approach, we generate *k-tuples* with $k \leq m$. A *k-tuples* $(x_1, \ldots, x_k)$ is a *partial solution* at level k if $P(x_i, x_j)$ is true for all $i \neq j \leq k$. The basic idea of the backtrack approach is to *extend* a partial solution at level k to one at level $k + 1$, and if this extension is impossible, then to go back to the partial solution at level $k - 1$ and attempt to generate a different partial solution at level k.

A nice way to organize the information inherent in the partial solutions is the backtrack search tree. Here, the empty partial solution () is taken as the root, and the partial solution $(x_1, \ldots, x_{k-1}, x_k)$ is represented as the child of the partial solution $(x_1, \ldots, x_{k-1})$. Following the computer science terminology, we often called a partial solution a *node*. It is often true that the computing cost of processing a node is independent of its level k. Under this assumption, the total computing cost of a search is equal to the number of nodes in the search tree times the cost of processing a node.

The number of nodes in the search tree can often be reduced by using the *symmetry* or *property-preserving* operations. If A is the incidence matrix of a projective plane, then the operations of permuting the rows and permuting the columns of A correspond only to reordering the lines and points of the plane. These operations preserve the property of being a projective plane. To see how they reduce the size of a search tree, consider the plane of order 10. There are $\binom{111}{11} \approx 4.7 \times 10^{14}$ choices for the first column, corresponding to the number of ways of placing 11 ones in the 111 rows. By using the row permutations, we can assume that all these ones are placed on the first 11 rows, reducing the number of partial solutions (x_1) from 4.7×10^{14} to 1. Mathematicians love to use the phrase "without loss of generality" to indicate a simplification by symmetry operations. So without loss of generality, the second column has only one choice—with a one in the first row and the remaining 10 ones in rows 12 to 21. Now, row 1 has nine remaining ones. By permuting columns, we can assume that these remaining ones of row 1 are in columns 3 to 11. Next, by row permutation, the remaining 10 ones of column 3 can be placed in rows 22 to 31. Continuing in this manner, it is not difficult to show that there is only one choice up to column 21. Beyond this point, symmetry operations are difficult to visualize, because they often involve combinations of row and column permutations.

This process of reducing the search tree by using symmetry operations is called *isomorph rejection*. Two partial solutions are said to be *isomorphic* if there is a symmetry operation mapping one to the other. To determine whether a solution

exists, it is only necessary to extend the non-isomorphic partial solutions, because if a partial solution can be extended to a complete solution, then every other isomorphic partial solution can also be so extended. The testing of whether two partial solutions are isomorphic is called *isomorphism testing*. The set of symmetry operations mapping a partial solution to itself forms an *automorphism group*. In the context of projective planes, an automorphism of a complete solution is also called a *collineation*.

The efficient use of the symmetry operations is one of the most difficult problems in a backtrack search. It is both a blessing and a curse—a blessing because there is always hope of further optimization and a curse because it is the source of many programming errors. However, following the method of MacWilliams et al., it is clear that the existence question of the plane of order 10 is headed towards a computer-based solution and that symmetry consideration is a necessary tool for efficiency reasons.

In his 1974 Ph.D. thesis, Carter picked up where MacWilliams et al. had left off. He showed that there are six possible starting configurations associated with a codeword of weight 16. By using the computer, he managed to eliminate four of the cases, as well as one subcase of the fifth. He used a total of about 100 hours of computer time, mostly on a CDC 6600 at the Institute for Defense Analyses (IDA) in Princeton, New Jersey, and also on a CDC 7600 at the Lawrence Radiation Laboratories in Berkeley, California. His search investigated about 3×10^8 nodes, or about 10^3 nodes per second.

This was the situation when we entered the picture. While we knew $w_{15} = 0$, the search for weight 16 codewords was about three-quarters done and the search for weight 12 codewords was presumed to be too difficult. The person who suggested the problem to us was John Thompson.

4. The home stretch. Around 1980, we purchased a share in a VAX-11/780 with the intention of running long mathematical programs during the off-hours when it was otherwise unused. In the same year, Thompson sent John McKay two papers [**28**, **29**] outlining a connection between a codeword of weight 12 and a set of fixed-point free involutions. When McKay showed me the papers, I was finally hooked. We had a computer looking for a problem and the search for codewords of weight 12 was a problem looking for a computer. I started by writing a simple computer program to play with the ideas. Little did I know that it would become a 9 year undertaking!

We realized very early that the weight 12 case was going to take a lot of computer time. So I showed my simple program to Larry Thiel, who has a reputation of making any computer program run faster. Right away, he saw ways of speeding it up. He and Stanley Swiercz wrote most of our computer programs.

We realized also that it is important to estimate how long a computer search would take. It would be useless to speed up a program by a factor of ten, if the resulting program still requires 100 years to run. However, speeding it up by a factor of 1000 would make the search feasible in our environment. By adapting Knuth's Monte Carlo method [**16**], we estimated that the search tree had about 4×10^{11} nodes. Hoping that we could process 10^5 nodes per second, we arrived at an estimate of 50 days of computer time.

I was going to present the final results at the Ninth Australian Conference on Combinatorial Mathematics in 1981, but the program development was slower

than expected. At the conference, I had to quickly change my plan and talked about a feasibility study of such a search [**17**]. The program was finally developed and the search finished late in 1982. We did not find a completed incidence matrix and so, $w_{12} = 0$ [**18**]. Ryser was happy with the result and encouraged us to continue. Hall was both excited and pessimistic. He wrote, "For the first time, I doubt that a plane of order 10 exists."

The search for weight 12 codewords took 183 days instead of the predicted 50. The main reason for the discrepancy was that we could process only about 3×10^4 nodes per second instead of 10^5. Yet, we learned a lot about the planning of the search, the estimation process, and the optimization techniques—knowledge which became useful in our later programs. In retrospect, we probably could have reduced the size of the search tree by another factor of 10 and come within the predicted time.

Even before the search for the weight 12 codeword was finished, we were looking for a way to solve the whole problem. One possibility is to determine w_{16} and then the weight enumerator. Unless the weight enumerator provides a contradiction, we still have to perform another search. Hall in [**14**] proved that if $w_{12} \leq 1211$, then there exist primitive weight 20 codewords, which intersect the lines of the plane in at most four points. We might be able to skip the step of finding w_{16} if we could go directly to weight 20. Unfortunately, we obtained an estimate of at least 10^{13} weight 20 starting configurations. Even though, as Hall remarked, it required only a few minutes of computer time to try each starting configuration, a few minutes times 10^{13} equaled 10^7 years—clearly beyond our capability. There was no short cut which could avoid computing the weight enumerator.

During my visit to the University of Cambridge in 1982, Thompson showed me Carter's thesis and suggested that we might be able to finish the remaining cases. It was the first time I had read his thesis and I was surprised by how familiar his computing techniques were. We had rediscovered a lot of what was already in his thesis! Thompson also made me promise to redo Carter's work on weight 16, just to have an independent verification of its correctness. Although I have yet to do it, I still intend to fulfil this promise.

In the mean time, Larry had developed an extremely useful prototyping program, called NPL. It had a very humble beginning. Working with incidence matrices meant that we often had to draw them and fill in large numbers of zeros and ones. As a labour saving move, we started drawing templates and then making photocopies. Larry took this one step further, printing the matrices with a computer rather than drawing the templates by hand. Soon we included the capability of initializing the starting configurations and later, the ability to backtrack was also included. Before long, it became a full-blown prototyping program, capable of performing estimations and adapting to new configurations. It turned out to be an extremely useful tool. We could quickly explore an idea for an "improvement" to the search. Ideas that turned out to increase the required CPU time were discarded. We could use it to plan our final search in detail, without having to write the program first. The most exciting stage was this planning stage.

Using NPL, we found a feasible way to solve the remaining cases of Carter. Compared to Carter's work, we found a better search order and a better use of the symmetry, both reducing the size of the search tree. Carefully optimised programs to implement the planned attack were written for each of the remaining cases.

Running on two VAX computers, it took the equivalent of 80 days of computing on a VAX-11/780 [**19**]. No completed incidence matrix was found and hence $w_{16} = 0$.

Now that w_{12}, w_{15} and w_{16} were all known to be zero, we could compute the weight enumerator of the binary error-correcting code for the plane of order 10 [**20**]. What catches the eye is that $w_{19} = 24{,}675$. If one follows the same method of finding starting configurations and trying to extend them, then one will either construct the plane or show that it does not exist.

Even before we finished our search for the weight 16 case, Thompson was already busy at work. In a letter dated October 24, 1983, he enclosed a copy of his unpublished note [**30**] with a list of 82 starting configurations for the weight 19 case. He constructed them by hand and had not checked for pairwise inequivalence. He ended the letter with "I hope you'll press on with this question, although perhaps I am too greedy."

It was a challenge that we could not refuse. I quickly wrote up a program and found 65 starting configurations. Larry also upgraded his NPL program to include isomorphism testing and found 66 configurations! After cross checking our results, I found that his numbers were correct. We also reconciled our numbers with those of Thompson. At last, we could see the finish line at a distance.

By simple counting arguments, 17 of the 66 cases can easily be eliminated. Four other cases are eliminated in a more ad hoc manner. Scott Crossfield, working with me as an NSERC Summer Undergraduate Assistant, obtained estimates for each of the remaining 45 cases. Our preliminary results indicated that the problem could be solved with another two years of computing time. These arguments and estimation results were presented in [**19**].

We decided to solve these cases by starting with the easy ones and hoping to discover better methods for the more difficult ones. We hedged our preliminary estimates by saying that it might be too optimistic. Our estimate of how fast the computer could test the acceptability of a column was based on our experience of the weight 12 and 16 cases. For weight 19, this test was much slower for technical reasons related to the fact that 19 is an odd number whereas the other two numbers are even. We could proceed only about 60 nodes per second, giving an estimate of 100 years to finish the search. To make the search feasible, we needed to speed it up by another factor of 100. The only possibility left was to use a faster computer, bringing to mind a supercomputer. Fortunately, our method could be easily adapted to use the vectoring capability of a supercomputer [**31**]. It was time to ask for help. Both Hall and Thompson suggested that IDA might be willing because it had helped Carter previously. After some discussion with Nick Patterson, the deputy director of the Communication Research Division at IDA, he agreed to run our yet undeveloped program on their CRAY-1A as the lowest priority job using up the otherwise idling computer. He and Douglas Wiedemann, who was on leave from IDA pursuing graduate studies at the University of Waterloo, gave us many suggestions on the efficient usage of the CRAY. In addition, Patterson looked after the day-to-day running of the program for over two years. He was the unsung hero in the successful completion of our work.

In the meantime, Concordia had acquired more and more VAX computers. Soon, we were running our program on five different computers. We figured that there was enough computing resources at Concordia to solve all but a few of the most difficult cases. So we targeted the CRAY program for only one type of

starting configuration which contained all the most difficult cases. The other configurations were left for our VAX computers to solve. All the cases at Concordia finished around January 1987 and took the equivalent of 800 days of VAX-11/780 computer time.

There are many considerations that go into developing a computer program which runs for months and years. Interruptions, ranging from power failures to hardware maintenance, are to be expected. A program should not have to restart from the very beginning for every interruption; otherwise, it may never finish. To solve the restarting problem, we partitioned our starting configurations into smaller cases and a message was output onto a log file every time one of them was finished. If there was an interruption, we just looked up the last completed case and continued from that point. For our programs, a convenient subdivision was at the "$A2$" boundary; more details can be found in [**19**, **21**]. The messages written on the log file also contained statistics about the run to allow comparison and verification of the results.

In order to avoid any unpleasant surprises involved in running programs at a long distance, we also put in a number of safety checks. Most of them were related to testing whether the sizes of the internal data structures were large enough. For lack of memory space, one could not merely allocate the largest expected size for every structure. Instructions were given on how to selectively increase the sizes. Some data structures were so specialized that their size could not be increased. We temporarily ignored the problem by deciding to handle them later, if necessary. Of course, we hoped that it would not be necessary.

We finished developing our CRAY program in 1986 and we were relieved to obtain a final estimate of about 3 months of CRAY computer time. It started running at IDA in the fall of 1986 and was to await completion in two to three years.

5. The finish line. On November 11, 1988, Patterson called and said that the run was finished. Always careful, he proposed to verify that there was a record of each subcase before sending the results on a magnetic tape back to us. Somehow, the news of the end of the plane of order 10 was spread around the combinatorics community and we were deluged with inquiries. On November 18, Patterson called again with some bad news. There was an error number 4 for one of the $A2$'s. What is an error number 4? It had been two years since we last looked at our CRAY program. Taking out an old listing, it took us a while to determine that error number 4 was a size problem for a data structure that could not be enlarged. This $A2$ could not be solved with the existing CRAY program! If it had been able to handle this $A2$, it would have taken about 30 seconds. The only other program around that could handle this case was our slow but adaptable NPL program. After further dividing this $A2$ into about 200 smaller subcases, it took NPL one day to solve one of them, giving an estimate of 200 days just to solve this $A2$! Meanwhile, Barry Cipra from *Science* called and said he would like to write an article about the non-existence of a plane of order 10. What should we tell everybody?

We decided that it was prudent to solve the problem first. We found that, with a small modification of the CRAY program, it could handle all but one of the subcases. So, we ran this offending subcase using NPL and bypassed the CRAY. The plane of order 10 was again dead on November 29, 1988.

There was a surprising amount of public interest in the non-existence of a finite projective plane of order 10. Besides Science, it was also reported in the *New York Times* and in *Scientific American*. One often asked question is, "How much computer time it took on the CRAY?" Unfortunately, we did not keep track of the computer time used on the CRAY, because we thought it was a useless figure other than demonstrating the difficulty of the problem. We were more interested in keeping statistics that might be useful for a future verification by someone else. So, we had to make an educated guess and said 3,000 hours. Later, Patterson suggested that it was probably closer to 2,000 hours. So, our CRAY program was proceeding nodes at a rate of about 2×10^4 nodes per second.

After the burst of publicity, we finally managed to read the magnetic tape containing the statistics. To our horror, we found another $A2$ with an error number 4. However, we knew exactly what to do this time and there was no panic. It was handled exactly the same way as the previous one. By the end of January 1989, the plane of order 10 was dead a third and, hopefully, the final time.

6. Is this really the end? This is not the first time that a computer has played an important role in "proving" a theorem. A notable earlier example is the four-color theorem [**3**]. Yet, these are not proofs in the traditional mathematical sense. It is impossible for any human being to check through all the calculations. From personal experience, it is extremely easy to make programming mistakes. We have taken many precautions, including the use of two different programs to cross check selective sample cases and the checking of internal consistency when isomorphism testing is performed. Yet, I want to emphasize that this is only an experimental result and it desperately needs an independent verification, or better still, a theoretical explanation.

There is, moreover, the possibility of an undetected hardware failure. A common error of this type is the random changing of bits in a computer memory, which could mean the loss of a branch of the search tree. This is the worst kind of hardware error, because we might lose solutions without realizing it. The CRAY-1A is reported to have such errors at the rate of about one per one thousand hours of computing. At this rate, we expect to encounter two to three errors! We did discover one such error by chance. After a hardware problem, Patterson reran the 1,000 $A2$'s just before the failure and the statistics have changed for the $A2$ processed just prior to the malfunction. How should one receive a "proof" that is almost guaranteed to contain several random errors?

Unfortunately, this is unavoidable in a computer-based proof—it is never absolute. However, despite this reservation, we argued in [**21**] that the possibility of hardware errors leading us to a wrong conclusion is extremely small. Since each $A2$ is in a separate run and there are about half a million non-isomorphic $A2$'s, the probability of one random hardware error affecting one specific $A2$ is about 2×10^{-6}. Suppose we accept that the weight enumerator is correct. Then if an undiscovered plane of order 10 exists, it would contain 24,675 weight 19 codewords. The 19 points in each such codeword give rise to an $A2$. Since we have searched through all non-isomorphic $A2$'s, we must have encountered these special $A2$'s. If all these 24,675 special $A2$'s are isomorphic, then only one out of about half a million non-isomorphic $A2$'s can be extended to the undiscovered plane. Even under this assumption, the probability of this special $A2$ being affected by two or three undetected hardware errors is less than 10^{-5}. Is it likely

that all 24,675 $A2$'s arising from an undiscovered plane are isomorphic? Since the plane is known to have a trivial collineation group [**1**, **15**, **36**], it is more likely that there are two or more non-isomorphic $A2$'s amongst the 24,675 cases. In this situation, the probability of hardware errors affecting all of them is infinitesimal. The same argument can be used even if we do not assume the correctness of previous computer-based results. Basically, the argument depends on the observation that if a plane of order 10 exists, then it can be constructed from many different starting points. Random hardware failures are unlikely to eliminate all of them. In other words, the fact that no one has yet constructed one is a very strong indication that it does not exist.

7. Epilogue. While we were tracing the origin of the existence problem of the plane of order 10, we talked to Dan Hughes, who has worked in this area for a long time and is famous for the Hughes planes which are named after him. He recounted the following story. In about 1957, at a Chinese restaurant in Chicago, Reinhold Baer, another mathematician well known for his work in group theory and projective planes, was trying to impress the younger Hughes by remarking that if the plane of order 10 was settled by a computer, he hoped not to be alive to see it. Baer got his wish but I do not think Herb Ryser shared this opinion. Ryser was happy that the weight 12 case was settled by a computer. I can only extrapolate and hope that he would also be happy that the whole problem has been "settled," even if by a computer.

REFERENCES

1. R. P. Anstee, M. Hall, Jr., and J. G. Thompson, Planes of order 10 do not have a collineation of order 5, *J. Comb. Theory, Series A*, 29 (1980) 39–58.
2. E. F. Assmus, Jr. and H. F. Mattson, Jr., On the Possibility of a Projective Plane of Order 10, Algebraic Theory of Codes II, Air Force Cambridge Research Laboratories Report AFCRL-71-0013, Sylvania Electronic Systems, Needham Heights, Mass., 1970.
3. K. Appel and W. Haken, Every planar map is four-colorable, *Bull. Amer. Math. Soc.*, 82 (1976) 711–712.
4. R. C. Bose, On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares, *Sankhyā*, 3 (1938) 323–338.
5. R. C. Bose and S. S. Shrikhande, On the falsity of Euler's conjecture about the non-existence of two orthogonal latin squares of order $4t + 2$, *Proc. Nat. Acad. Sci.*, 45 (1959) 734–737.
6. R. C. Bose, S. S. Shrikhande, and E. T. Parker, Further results on the construction of mutually orthogonal latin squares and the falsity of Euler's conjecture, *Canad. J. Math.*, 12 (1960) 189–203.
7. R. H. Bruck and H. J. Ryser, The non-existence of certain finite projective planes, *Can. J. Math.*, 1 (1949) 88–93.
8. A. Bruen and J. C. Fisher, Blocking sets, k-arcs and nets of order ten, *Advances in Math.* 10 (1973) 317–320.
9. J. L. Carter, On the Existence of a Projective Plane of Order Ten, Ph.D. thesis, Univ. of Calif., Berkeley, 1974.
10. S. Chowla and H. J. Ryser, Combinatorial problems, *Can. J. Math.*, 2 (1950) 93–99.
11. R. H. F. Denniston, Non-existence of a certain projective plane, *J. Austral. Math. Soc.*, 10 (1969) 214–218.
12. L. Euler, Recherches sur une nouvelle espèce de quarrés magiques, *Verh. Zeeuwsch. Genootsch. Wetensch. Vlissengen*, 9 (1782) 85–239.
13. M. Hall, Jr. and H. J. Ryser, Normal completions of incidence matrices, *Amer. J. Math*, 76 (1954) 581–589.
14. M. Hall, Jr., Configurations in a plane of order 10, *Ann. Discrete Math.*, 6 (1980) 157–174.
15. Z. Janko and T. van Trung, Projective planes of order 10 do not have a collineation of order 3, *J. Reine Angew. Math.*, 325 (1981) 189–209.

16. D. E. Knuth, Estimating the efficiency of backtrack programs, *Mathematics of Computations*, 29 (1975) 121–136.
17. C. W. H. Lam, L. Thiel, and S. Swiercz, A feasibility study of a search for ovals in a projective plane of order 10, *Proceeding of the Ninth Australian Conference on Combinatorial Mathematics*, Springer-Verlag Lecture Notes in Mathematics, 952 (1982) 349–352.
18. C. W. H. Lam, L. Thiel, S. Swiercz, and J. McKay, The nonexistence of ovals in a projective plane of order 10, *Discrete Mathematics*, 45 (1983) 319–321.
19. C. Lam, S. Crossfield, and L. Thiel, Estimates of a computer search for a projective plane of order 10, *Congressus Numerantium*, 48 (1985) 253–263.
20. C. W. H. Lam, L. Thiel, and S. Swiercz, The nonexistence of code words of weight 16 in a projective plane of order 10, *J. of Combinatorial Theory*, Series A 42 (1986) 207–214.
21. C. W. H. Lam, L. H. Thiel, and S. Swiercz, The non-existence of finite projective planes of order 10, *Can. J. Math.*, XLI (1989) 1117–1123.
22. J. MacWilliams, N. J. A. Sloane, and J. G. Thompson, On the existence of a projective plane of order 10, *J. Combinatorial Theory*, Sec. A. 14 (1973) 66–78.
23. E. T. Parker, Construction of some sets of mutually orthogonal Latin squares, *Proc. Amer. Math. Soc.*, 10 (1959) 946–949.
24. _____, Orthogonal Latin squares, *Proc. Nat. Acad. Sci. U.S.A.*, 45 (1959) 859–862.
25. H. J. Ryser, A note on a combinatorial problem, *Proc. Amer. Math. Soc.*, 1 (1950), 422–424.
26. _____, Combinatorial Mathematics, The Carus Mathematical Monographs, Math. Assoc. of America, 1963.
27. G. Tarry, Le problème des 36 officiers, *C. R. Assoc. Fran. Av. Sci.*, 1 (1900) 122–123, Vol. 2 (1901), 170–203.
28. J. G. Thompson, Fixed Point Free Involutions and Finite Projective Planes, Durham Conference on Finite Groups, ed. M. Collins, 1978.
29. _____, Ovals in a projective plane of order 10, unpublished.
30. _____, Extremal 19-sets in the F_2-code of a projective plane of order 10, unpublished.
31. L. H. Thiel, C. Lam, and S. Swiercz, Using a CRAY-1 to perform backtrack search, Proc. of the Second International Conference on Supercomputing, San Francisco, III (1987) 92–99.
32. O. Veblen, A system of axioms for geometry, *Trans. Amer. Math. Soc.*, 5 (1904) 343–384.
33. O. Veblen and W. H. Bussey, Finite projective geometries, *Trans. Amer. Math. Soc.*, 7 (1906) 241–259.
34. O. Veblen and J. H. M. Wedderburn, Non-Desarguesian and non-Pascalian geometries, *Trans. Amer. Math. Soc.*, 8 (1907) 379–388.
35. R. J. Walker, An enumerative technique for a class of combinatorial problems, *Proc. AMS Symp. Appl. Math.*, X (1960) 91–94.
36. S. H. Whitesides, Collineations of projective planes of order 10, parts I and II, *J. Comb. Theory, Series A*, 26 (1979) 249–277.

The Fiftieth William Lowell Putnam Mathematical Competition

LEONARD F. KLOSINSKI, *Santa Clara University*
GERALD L. ALEXANDERSON, *Santa Clara University*
LOREN C. LARSON, *St. Olaf College*

The following results of the fiftieth William Lowell Putnam Mathematical Competition, held on December 2, 1989, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, $5,000, was awarded to the Department of Mathematics of Harvard University. The members of the winning team were: Jeremy A. Kahn, Raymond M. Sidney, and Eric K. Wepsic; each was awarded a prize of $250.

The second prize, $2,500, was awarded to the Department of Mathematics of Princeton University. The members of the winning team were: David J. Grabiner, Mathew D. Mullin, and Rahul V. Pandharipande; each was awarded a prize of $200.

The third prize, $1,500, was awarded to the Department of Mathematics of the University of Waterloo. The members of the winning team were: Grayden Hazenberg, Stephen M. Smith, and Colin M. Springer; each was awarded a prize of $150.

The fourth prize, $1,000, was awarded to the Department of Mathematics of Yale University. The members of the winning team were: Bruce E. Kaskel, Andrew H. Kresch, and Sihao Wu; each was awarded a prize of $100.

The fifth prize, $500, was awarded to the Department of Mathematics of Rice University. The members of the winning team were: Hubert L. Bray, John W. McIntosh, and David S. Metzler; each was awarded a prize of $50.

The six highest ranking individual contestants, in alphabetical order, were Christo Athanasiadis, Massachusetts Institute of Technology; William P. Cross, California Institute of Technology; Andrew H. Kresch, Yale University; Colin M. Springer, University of Waterloo; Ravi D. Vakil, University of Toronto; and Sihao Wu, Yale University. Each of these was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of $500 by the Putnam Prize Fund.

The next four highest ranking individuals, in alphabetical order, were William Chen, Washington University, St. Louis; Jordan S. Ellenberg, Harvard University; David J. Grabiner, Princeton University; and Raymond M. Sidney, Harvard University. Each was awarded a prize of $250.

The following teams, named in alphabetical order, received honorable mention: California Institute of Technology, with team members William P. Cross, Robert Southworth, and Glenn Tesler; the University of California, Berkeley, with team members Warren M. Lam, Jordan Lampe, and Jonathan C. Ruyle; Carnegie Mellon University, with team members Petros I. Hadjicostas, Eric Lauer, and Karl M. Westerberg; the University of Chicago, with team members Justin Boyan, Doyle Tanner, and Andrew Yeh; and Stanford University, with team members John C. Loftin, Greg G. Martin, and Christopher S. Wilson.

Honorable mention was achieved by the following thirty-five individuals named in alphabetical order: Edgar D. Bailey, Jr., University of Texas, Austin; Daniel J. Bernstein, New York University; Hubert L. Bray, Rice University; Nicholai I. Chaudarov, Brandeis University; Andrew Chou, Massachusetts Institute of Technology; Michael Cole, Hofstra University; Matthew M. Cook, University of Illinois, Champaign-Urbana; Daniel P. Cory, Stanford University; Kevin B. Ford, California State University, Chico; Michael L. Hutchings, Harvard University; Jeremy A. Kahn, Harvard University; Jordan Lampe, University of California, Berkeley; Gregory D. Landweber, Princeton University; Daniel D. Lee, Harvard University; Daniel I. Lieberman, Princeton University; Sanjoy S. Mahajan, Stanford University; Russell A. Manning, California Institute of Technology; David M. Maymudes, Harvard University; John W. McIntosh, Rice University; Christopher J. Monsour, University of Maryland, College Park; Rahul Vijay Pandharipande, Princeton University; Eric M. Rains, Case Western Reserve University; Stephen A. Smith, University of Waterloo; Steven M. Stadnicki, Clarkson University; Andras Vasy, Stanford University; Eric H. Veach, University of Waterloo; Nikolai I. Weaver, Harvey Mudd College; Eric K. Wepsic, Harvard University; Karl M. Westerberg, Carnegie Mellon University; David B. Wilson, Massachusetts Institute of Technology; Michael P. Wolf, Harvard University; Ali F. Yegulalp, Columbia University; Andrew Yeh, University of Chicago; Zhaoliang Zhu, Yale University; and Michael E. Zieve, Harvard University.

The other individuals who achieved ranks among the top 94, in alphabetical order of their schools, were: University of Alberta, Graham C. Denham, Brendan M. Mumey; University of British Columbia, Wayne J. Broughton, Gregory F. Wellman; Brown University, Martin M. Wattenberg; California Institute of Technology, Tien-Yee Chiu, Michael G. Greenblatt, Gwoho H. Liu, Robert G. Southworth; University of California, Berkeley, Bryan F. Clair, Wee-Lang Heng; University of California, Santa Barbara, John C. Carey; University of California, San Diego, Franz J. Wrasidlo; University of Chicago, Adrian Tanner; Duke University, Will A. Schneeberger; Emory University, Charles D. McDonell; Georgia Institute of Technology, Jeffrey W. Herrmann; Harvard University, J. Stewart Burns, Roland B. Dreier, Daniel Eric Gottesman, F. Dean Hildebrandt, Michael C. Jablecki, Michael D. Mitzenmacher, Matthew C. Weiner; University of Maryland, College Park, Eric M. Boesch, Jay P. Chawla; Massachusetts Institute of Technology, Todd W. Rowland; Michigan State University, Jacob R. Lorch; Princeton University, Timothy Y. Chow, Paul E. Ericksen, William P. Minicozzi, Matthew D. Mullin; Reed College, Nathaniel P. Thurston; Rice University, John Kenneth Burton, Jr., David S. Metzler; Simon Fraser University, Richard S. Kiss; Stanford University, Thomas H. Chung, John C. Loftin, Greg G. Martin; University of Texas, Austin, Bryan W. Taylor; Trinity College, Hartford, Marshall A. Whittlesey; Washington University, St. Louis, Adam M. Costello, Jordan A. Samuels, Peter S. Shawhan; University of Waterloo, Graydon H. Hazenberg, Marc S. Ordower, Giuseppe Russo; Wellesley College, Yihao L. Zhang; and the University of Wisconsin, Madison, Cavan C. Fang.

There were 2392 individual contestants from 373 colleges and universities in Canada and the United States in the competition of December 2, 1989. Teams were entered by 288 institutions.

The Questions Committee for the fiftieth competition consisted of Gerald A. Heuer (Chair), Paul R. Halmos, and Kenneth A. Stolarsky; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1. How many primes among the positive integers, written as usual in base 10, are such that their digits are alternating 1's and 0's, beginning and ending with 1?

Problem A-2. Evaluate $\int_0^a \int_0^b e^{\max\{b^2x^2, a^2y^2\}}\, dy\, dx$, where a and b are positive.

Problem A-3. Prove that if

$$11z^{10} + 10iz^9 + 10iz - 11 = 0,$$

then $|z| = 1$. (Here z is a complex number and $i^2 = -1$.)

Problem A-4. If α is an irrational number, $0 < \alpha < 1$, is there a finite game with an honest coin such that the probability of one player winning the game is α? (An honest coin is one for which the probability of heads and the probability of tails are both $1/2$. A game is finite if with probability 1 it must end in a finite number of moves.)

Problem A-5. Let m be a positive integer and let $\mathscr{S}$ be a regular $(2m + 1)$-gon inscribed in the unit circle. Show that there is a positive constant A, independent of m, with the following property. For any point p inside $\mathscr{S}$ there are two distinct vertices v_1 and v_2 of $\mathscr{S}$ such that

$$\left| |p - v_1| - |p - v_2| \right| < \frac{1}{m} - \frac{A}{m^3}.$$

Here $|s - t|$ denotes the distance between the points s and t.

Problem A-6. Let $\alpha = 1 + a_1x + a_2x^2 + \cdots$ be a formal power series with coefficients in the field of two elements. Let

$$a_n = \begin{cases} & \text{if every block of zeros in the binary} \\ 1 & \text{expansion of } n \text{ has an even number} \\ & \text{of zeros in the block,} \\ 0 & \text{otherwise.} \end{cases}$$

(For example, $a_{36} = 1$ because $36 = 100100_2$, and $a_{20} = 0$ because $20 = 10100_2$.) Prove that $\alpha^3 + x\alpha + 1 = 0$.

Problem B-1. A dart, thrown at random, hits a square target. Assuming that any two parts of the target of equal area are equally likely to be hit, find the probability that the point hit is nearer to the center than to any edge. Express your answer in the form $(a\sqrt{b} + c)/d$, where a, b, c, d are positive integers.

Problem B-2. Let S be non-empty set with an associative operation that is left and right cancellative ($xy = xz$ implies $y = z$, and $yx = zx$ implies $y = z$). Assume that for every a in S the set $\{a^n\colon n = 1, 2, 3, \ldots\}$ is finite. Must S be a group?

Problem B-3. Let f be a function on $[0, \infty)$, differentiable and satisfying

$$f'(x) = -3f(x) + 6f(2x)$$

for $x > 0$. Assume that $|f(x)| \leqslant e^{-\sqrt{x}}$ for $x \geqslant 0$ (so that $f(x)$ tends rapidly to 0 as x increases). For n a non-negative integer, define

$$\mu_n = \int_0^\infty x^n f(x)\, dx$$

(sometimes called the nth moment of f).

a. Express μ_n in terms of μ_0.

b. Prove that the sequence $\{\mu_n(3^n/n!)\}$ always converges, and that the limit is 0 only if $\mu_0 = 0$.

Problem B-4. Can a countably infinite set have an uncountable collection of non-empty subsets such that the intersection of any two of them is finite?

Problem B-5. Label the vertices of a trapezoid T (quadrilateral with two parallel sides) inscribed in the unit circle as A, B, C, D so that AB is parallel to CD and A, B, C, D are in counterclockwise order. Let s_1, s_2 and d denote the lengths of the line segments AB, CD, and OE, where E is the point of intersection of the diagonals of T, and O is the center of the circle. Determine the least upper bound of $(s_1 - s_2)/d$ over all such T for which $d \neq 0$, and describe all cases, if any, in which it is attained.

Problem B-6. Let $(x_1, x_2, \ldots, x_n)$ be a point chosen at random from the n-dimensional region defined by $0 < x_1 < x_2 < \cdots < x_n < 1$. Let f be a continuous function on $[0, 1]$ with $f(1) = 0$. Set $x_0 = 0$ and $x_{n+1} = 1$. Show that the expected value of the Riemann sum

$$\sum_{i=0}^{n} (x_{i+1} - x_i) f(x_{i+1})$$

is $\int_0^1 f(t)P(t)\, dt$, where P is a polynomial of degree n, independent of f, with $0 \leqslant P(t) \leqslant 1$ for $0 \leqslant t \leqslant 1$.

SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geqslant i \geqslant 0$ is the number of students among the top 199 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1 (92, 2, 6, 7, 0, 0, 0, 9, 2, 8, 41, 32)

Solution. One. Let N_k be the number

$$101010\ldots101$$

with exactly k digits equal to 0. If $k = 1$, so that $N_k = 101$, then N_k is prime. All other N_k's are composite, as the following reasoning shows. If k is odd, then 101 divides N_k. If k is even, then

$$11N_k = R \cdot S,$$

where R is the number consisting of $k + 1$ digits all equal to 1 and S is the

number with $k + 2$ digits beginning and ending with 1 and having only 0's in between. One of the numbers R and S divides N_k.

A-2 (141, 6, 29, 0, 0, 0, 0, 0, 5, 7, 4, 7)

Solution. Divide the region into two parts by the diagonal line $ay = bx$ to get

$$\begin{aligned}\int_0^a \int_0^b e^{\max\{b^2x^2, a^2y^2\}}\,dy\,dx &= \int_0^a \int_0^{bx/a} e^{b^2x^2}\,dy\,dx + \int_0^b \int_0^{ay/b} e^{a^2y^2}\,dx\,dy \\ &= \int_0^a \frac{bx}{a} e^{b^2x^2}\,dx + \int_0^b \frac{ay}{b} e^{a^2y^2}\,dy \\ &= \frac{e^{a^2b^2} - 1}{ab}.\end{aligned}$$

A-3 (13, 4, 4, 2, 0, 0, 0, 0, 6, 8, 56, 106)

Solution. We have $z^9 = \dfrac{11 - 10iz}{11z + 10i}$. If $z = a + bi$, then

$$|z^9| = \left|\frac{11 - 10iz}{11z + 10i}\right| = \sqrt{\frac{|11^2 + 220b + 10^2(a^2 + b^2)|}{|11^2(a^2 + b^2) + 220b + 10^2|}} \equiv \frac{f(a,b)}{g(a,b)}.$$

If $a^2 + b^2 > 1$, then $g(a, b) > f(a, b)$, making $|z^9| < 1$, a contradiction. If $a^2 + b^2 < 1$, then $f(a, b) > g(a, b)$, making $|z^9| > 1$, a contradiction. Thus, $|z| = 1$.

Second Solution. Rouché's Theorem states that if $|g(z) - f(z)| < |f(z)|$ for z on $|z| = \alpha$ then g and f have the same number of zeros inside $|z| = \alpha$. Let $g(z) = 11z^{10} + 10iz^9 + 10iz - 11$ and $f(z) = 10iz - 11$. Then $|g(z) - f(z)| = |11z^{10} + 10iz^9| = |z|^9|i||11z + 10i| < |10iz - 11|$ if $|z| < 1$. Now $f(z)$ has its only root at $11/10i$ so that $g(z)$ has no zeros inside $|z| < 1$. If z is a zero of g and $|z| > 1$, then $-1/z$ is also a zero of g (which we've seen doesn't happen), and therefore, the only zeros of g are on the circle $|z| = 1$.

A-4 (54, 22, 16, 4, 0, 0, 0, 0, 4, 2, 53, 44)

Solution. Yes. Write the number

$$\alpha = \sum_{n=1}^{\infty} \frac{\alpha_n}{2^n}$$

in its binary representation,

$$\alpha_n = 0 \text{ or } 1, \qquad n = 1, 2, 3, \ldots,$$

and play the game as follows. Keep tossing the coin, recording the result, say β_n, of the nth toss as 0 when it falls tails and 1 when it falls heads; you win the game if the first time that $\beta_n \neq \alpha_n$ your result, β_n, is 0 (and, therefore, $\alpha_n = 1$). In other words, you win if at the first time when the β sequence differs from the α sequence, the β sequence is smaller.

The motivation for this procedure is to think of α as determining the interval $(0,\alpha)$, and think of the coin tosses as determining a number in $[0, 1]$; "win" means "land in the prescribed interval". There are many other solutions; this might be one of the simplest.

How can you win? To see the answer, it helps to introduce some notation: let $n_1, n_2, n_3, \ldots$ be the positions of the 1's that occur in the sequence, α. One way to win is to toss so that the β sequence agrees with the α sequence for the first $n_1 - 1$ terms and then disagrees. The probability of that is $1/2$ to the power n_1, which is equal to the partial sum up to n_1 of the binary representation of α. The next way to win is to toss so that the β sequence agrees with the α sequence for the first $n_2 - 1$ terms and then disagrees. The probability of that is $1/2$ to the n_2, which is equal to the number obtained from the partial sum up to n_2 of the binary representation of α by replacing the first 1 with a 0. These two ways of winning are the beginning of an infinite sequence. The (infinite) sum of the corresponding probabilities is exactly the binary representation of α.

A-5 (6, 1, 3, 0, 0, 0, 0, 1, 0, 2, 49, 137)

Solution. The diameter of $\mathscr{S}$ is $2\cos(\pi/2(2m+1))$, so all the $|p - v_i|$ for $1 \leqslant i \leqslant 2m+1$ fall into the interval $[0, 2\cos(\pi/2(2m+1))]$. Hence some two of them differ by at most

$$\frac{2\cos\left(\dfrac{\pi}{2(2m+1)}\right)}{2m}.$$

Let

$$f(m) = \frac{1 - \cos\left(\dfrac{\pi}{2(2m+1)}\right)}{\left(\dfrac{1}{m}\right)^2}.$$

This is a positive term sequence with limit $\pi^2/32$ and therefore this sequence has a positive minimum value. Therefore,

$$\frac{1 - \cos\left(\dfrac{\pi}{2(2m+1)}\right)}{\left(\dfrac{1}{m}\right)^2} > A$$

for some $A > 0$. Hence

$$\cos\left(\frac{\pi}{2(2m+1)}\right) < 1 - \frac{A}{m^2}$$

and the result follows.

A-6 (3, 1, 0, 0, 0, 0, 0, 1, 0, 4, 49, 141)

Solution. We show that $\alpha^4 + x\alpha^2 + \alpha = 0$.
Note that

$$\alpha^2 = \left(\sum_{n=0}^{\infty} a_n x^n\right)^2 = \sum_{n=0}^{\infty} a_n x^{2n}, \qquad \alpha^4 = \sum_{n=0}^{\infty} a_n x^{4n}, \qquad x\alpha^2 = \sum_{n=0}^{\infty} a_n x^{2n+1}.$$

Therefore,

$$\alpha^4 + x\alpha^2 + \alpha = \sum_{\substack{n=0 \\ 2\mid n,\, 4\nmid n}}^{\infty} a_n x^n + \sum_{\substack{n=0 \\ 4\mid n}}^{\infty} (a_{n/4} + a_n) x^n + \sum_{\substack{n=0 \\ n \text{ odd}}}^{\infty} (a_n + a_{(n-1)/2}) x^n.$$

If $2|n, 4|n$, then $a_n = 0$. If $4|n$ then $a_{n/4} = a_n$, so $a_{n/4} + a_n = 0$. If n is odd, then $a_{(n-1)/2} = a_n$, so $a_n + a_{(n-1)/2} = 0$.

B-1 (144, 10, 0, 0, 0, 0, 0, 0, 0, 0, 42, 3)

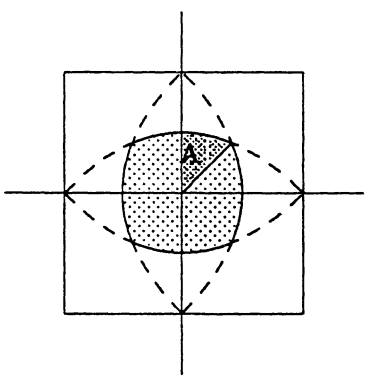
Solution. Consider a 2×2 dartboard centered at the origin with vertices at $(1, 1)$, $(-1, 1)$, $(-1, -1)$, and $(1, -1)$. A point (x, y) in the square is closer to the center of the board than the top edge of the board if and only if

$$\sqrt{x^2 + y^2} \leqslant 1 - y,$$

or equivalently,

$$y \leqslant \frac{1 - x^2}{2}.$$

A similar parabolic region arises for each of the other three sides, and therefore the region of points closer to the center than any edge has the shape of the shaded region in the following figure.



Let A denote the area of this region in the first quadrant bounded by $y = (1 - x^2)/2$ and $y = x$. Then, the probability we desire is

$$\frac{8 \text{ Area } A}{4} = 2 \text{ Area } A = 2\int_0^{\sqrt{2}-1}\left(\frac{1 - x^2}{2} - x\right) dx = \frac{4\sqrt{2} - 5}{3}.$$

B-2 (49, 44, 17, 10, 0, 0, 0, 0, 3, 4, 41, 31)

Solution. Yes. Let a be an arbitrary element of S. The set $\{a^n: n = 1, 2, 3, \ldots\}$ is finite, and therefore $a^m = a^n$ for some m, n with $m > n \geqslant 1$. By cancellation we have $a^{r(a)} = a$, where $r(a) = m - n + 1 > 1$. If x is any element of S, then $aa^{r(a)-1}x = a^{r(a)}x = ax$, and this implies that $a^{r(a)-1}x = x$. Similarly, we see that $xa^{r(a)-1} = x$, and the element $e \equiv a^{r(a)-1}$ is an identity. The identity element is unique, for if e' is another identity, then $e = ee' = e'$. If $r(a) > 2$ then $a^{r(a)-2}$ is an inverse of a, and if $r(a) = 2$ then $a^2 = a = e$ and a is its own inverse. Thus S is a group.

B-3 (36, 4, 18, 48, 4, 10, 3, 1, 4, 1, 9, 61)

Solution. a. Clearly, $\int_0^\infty x^n g(x)\,dx$ exists for $g(x)$ equal to $f(x)$ or $f'(x)$ because $f(x)$ tends rapidly to 0. Hence,

$$\int_0^\infty x^n f'(x)\,dx = -3\mu_n + 6\int_0^\infty x^n f(2x)\,dx.$$

Using parts on the integral on the left and the substitution $u = 2x$ on the integral on the right, we obtain

$$x^n f(x)\Big]_0^\infty - n\mu_{n-1} = -3\mu_n + \frac{6}{2^{n+1}}\mu_n.$$

Since $x^n f(x) \to 0$ as $x \to \infty$ for any $n > 0$ we have

$$\mu_n = \frac{n}{3}\,\frac{1}{1 - \frac{1}{2^n}}\,\mu_{n-1}.$$

Iteration now yields

$$\mu_n = \frac{n!}{3^n}\,\frac{1}{\prod_{m=1}^{n}\left(1 - \frac{1}{2^m}\right)}\,\mu_0.$$

b. Since $\Sigma_{m=1}^\infty (1/2^m) < \infty$, the infinite product $\Pi_{m=1}^\infty (1 - (1/2^m))$ converges to a nonzero finite limit and the result follows.

B-4 (52, 7, 0, 0, 0, 0, 0, 0, 0, 1, 62, 77)

Solution. Yes. Consider the countably infinite set Q of rational numbers in $[0, 1]$. For each irrational number α in $[0, 1]$, let Q_α be a subset of Q that has α as its unique cluster point. There are uncountably many Q_α's, each of which is countably infinite (and, in particular, not empty). If $Q_\alpha \cap Q_\beta$ is infinite, then it clusters at both α and β, and therefore $\alpha = \beta$.

Second Solution. Let L denote the set of lattice points in the plane: $\{(m, n) | m,n \in Z\}$. Let S_θ be the set of points in L at a distance less than or equal to one from the line $y = \theta x$, $-1 \leqslant \theta \leqslant 1$. Then $|S_{\theta_1} \cap S_{\theta_2}|$ is a finite set if and only if $\theta_1 \neq \theta_2$.

B-5 (17, 17, 1, 1, 0, 0, 0, 1, 17, 29, 37, 79)

Solution. Choose coordinates so that AB and CD are vertical, and so that E lies on the x-axis. For an appropriate slope m we see that B and D lie on the intersection of $y = m(x + d)$ with $x^2 + y^2 = 1$. Thus,

$$y^2 + \left(\frac{y}{m} - d\right)^2 = 1$$

$$\left(1 + \frac{1}{m^2}\right)y^2 - \frac{2d}{m}y + (d^2 - 1) = 0$$

$$y^2 - \left(\frac{2dm}{1 + m^2}\right)y + \frac{d^2 - 1}{1 + m^{-2}} = 0.$$

Now (consider y coordinates) $s_1 - s_2$ is *twice* the sum of the roots of this equation, and therefore,

$$s_1 - s_2 = 2\left(\frac{2m}{1+m^2}\right)d \leqslant 2d,$$

with equality if and only if $m = 1$. Thus for $d \neq 0$ the maximum value of $(s_1 - s_2)/d$ is 2 and it is attained if and only if the diagonals intersect at right angles and T is not a square (since $d = 0$ for a square).

B-6 (0, 1, 1, 1, 0, 0, 0, 2, 0, 11, 35, 148)

Solution. The volume in R^n of all points $(x_1, x_2, \ldots, x_n)$ with $0 < x_1 < x_2 < \cdots < x_n < 1$ is

$$\int_0^1 \int_0^{x_n} \cdots \int_0^{x_2} dx_1\, dx_2 \ldots dx_n = \frac{1}{n!}.$$

Thus, we need to divide

$$M = \sum_{i=0}^{n} \int_0^1 \int_0^{x_n} \cdots \int_0^{x_2} (x_{i+1} - x_i) f(x_{i+1})\, dx_1 \ldots dx_n$$

by $1/n!$. When $0 \leqslant i \leqslant n-1$ iterated integration yields for a typical summand

$$\int_0^1 \int_0^{x_n} \cdots \int_0^{x_{i+1}} (x_{i+1} - x_i) \frac{x_i^{i-1}}{(i-1)!} f(x_{i+1})\, dx_i \ldots dx_n$$

$$\int_0^1 \int_0^{x_n} \cdots \int_0^{x_{i+2}} \frac{x_{i+1}^{i+1}}{(i+1)!} f(x_{i+1})\, dx_{i+1} \ldots dx_n$$

$$\int_0^1 \int_{x_{i+1}}^1 \int_0^{x_n} \cdots \int_{x_{i+1}}^{x_{i+3}} \frac{x_{i+1}^{i+1}}{(i+1)!} f(x_{i+1})\, dx_{i+2} \ldots dx_n\, dx_{i+1}$$

$$\int_0^1 \frac{(1 - x_{i+1})^{n-(i+1)}}{(n-(i+1))!} \frac{x_{i+1}^{i+1}}{(i+1)!} f(x_{i+1})\, dx_{i+1}.$$

Multiplication by $n!$ reveals that the kernels are "almost" the terms of the binomial expansion of $((1 - x_{i+1}) + x_{i+1})^n$. In fact by adding in and subtracting out the $i = -1$ term we get

$$n!M = \int_0^1 f(t)\left(1 - (1-t)^n\right) dt + n!J_n,$$

where

$$\begin{aligned} J_n &= \int_0^1 \int_0^{x_n} \cdots \int_0^{x_2} (1 - x_n) f(1)\, dx_1 \ldots dx_n \\ &= \int_0^1 \int_0^{x_n} (1 - x_n) \frac{x_{n-1}^{n-2}}{(n-2)!} f(1)\, dx_{n-1}\, dx_n \\ &= \frac{1}{(n+1)!} f(1), \end{aligned}$$

so the expected value is

$$\int_0^1 f(t) P_n(t)\, dt + \frac{1}{n+1} f(1)$$

where $P_n(t) = 1 - (1-t)^n$.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Symmetry Classes: Functions of Three Variables

N. METROPOLIS
Los Alamos National Laboratory, Los Alamos, NM 87545

GIAN-CARLO ROTA*
Los Alamos National Laboratory and
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

I. Introduction. It is well known that every function of two variables $f(x, y)$ is uniquely expressible as the sum of a symmetric function f_s and a skew-symmetric function f_a, defined, respectively, as

$$f_s(x, y) = \frac{f(x, y) + f(y, x)}{2} \qquad f_a(x, y) = \frac{f(x, y) - f(y, x)}{2}.$$

The attempt to generalize this fact to functions of n variables led Alfred Young to develop his theory of symmetrizers, now subsumed in the theory of representations of the symmetric group.

According to this theory, a function of n variables is uniquely expressible as the sum of p_n functions, each one belonging to a different symmetry class, where p_n is the number of partitions of the integer n. Unfortunately, a simple intuitive description of such symmetry classes has never been given except for $n = 2$.

It is known that for functions of three variables, there is only one other symmetry class besides the two obvious symmetry classes of symmetric functions and of skew-symmetric functions. We give this third symmetry class a very simple characterization, one that seems to have been overlooked. We show that it consists of all *cyclic-symmetric* functions. We prove that every function of three variables is uniquely expressible as the sum of a symmetric function, a skew-symmetric function and a cyclic-symmetric function.

To make this note self-contained, we have added a short derivation of some known formulas.

We believe that the underlying idea of this note will extend to functions of n variables. We hope the present note will at least entice the reader to further study of the vast theory of symmetry classes.

II. Symmetry Classes. We shall deal with everywhere defined real-valued functions of three variables taken from a set $\{x, y, z\}$. The set of all such functions forms a vector space V over the field of real numbers. This setup is deliberately limited in scope for purposes of clarity.

If σ is a permutation on the set $\{x, y, z\}$ mapping x to $x\sigma$ and if f is a function, we write

$$\sigma f(x, y, z) = f(x\sigma, y\sigma, z\sigma).$$

*Partially supported by NSF Grant No. MCS 8104855.

A subspace W of the vector space V will be said to be a *symmetry class* if whenever $f \in W$, then $\sigma f \in W$.

Our objective is to give a simple description of all symmetry classes.

There are three notable symmetry classes, two of which are well known:

(1) The symmetry class S consisting of all *symmetric functions*, i.e., all functions such that $f(x, y, z) = f(x\sigma, y\sigma, z\sigma)$ for all permutations σ of the set $\{x, y, z\}$;

(2) The symmetry class A of all *skew-symmetric* functions, i.e., all functions such that $f(x, y, z) = (\text{sign } \sigma)f(x\sigma, y\sigma, z\sigma)$ for all permutations σ of the set $\{x, y, z\}$. Here, $(\text{sign } \sigma)$ is the signature of the permutation σ. Since we are dealing with three variables, $(\text{sign } \sigma)$ will be -1 if σ is a transposition of two variables, and $(\text{sign } \sigma)$ will be 1 if σ is the identity or else if σ is a cyclic permutation.

(3) The third is the symmetry class C of *cyclic-symmetric* functions, defined as the subspace C of all functions f satisfying the two equations

$$\begin{aligned} f(x, y, z) + f(z, x, y) + f(y, z, x) &= 0 \\ f(x, z, y) + f(y, x, z) + f(z, y, x) &= 0. \end{aligned} \tag{*}$$

The preceding definition makes sense:

PROPOSITION. *The subspace C of V consisting of all functions satisfying (*) is a symmetry class.*

Proof. It suffices to show that if f is cyclic-symmetric and σ is a transposition of the set $\{x, y, z\}$, then σf is cyclic-symmetric. Now, by definition, the function σf is cyclic-symmetric if and only if it satisfies the equations

$$\begin{aligned} \sigma f(x, y, z) + \sigma f(z, x, y) + \sigma f(y, z, x) &= 0 \\ \sigma f(x, z, y) + \sigma f(y, x, z) + \sigma f(z, y, x) &= 0. \end{aligned} \tag{**}$$

For the sake of the argument, suppose σ is a transposition of x and z. The above equations (**) then reduce to the pair of equations

$$\begin{aligned} f(z, y, x) + f(x, z, y) + f(y, x, z) &= 0 \\ f(z, x, y) + f(y, z, x) + f(x, y, z) &= 0. \end{aligned}$$

But note that these equations are identical except for order to the two equations (*).

The same argument works for any transposition. Since any permutation is the product of transpositions, we conclude that σf is cyclic-symmetric for any permutation σ, as desired. □

It is easy to give examples of cyclic-symmetric functions. If $f(x, y, z)$ is any function, then the function $g(x, y, z) = f(x, y, z) - f(z, x, y)$ is cyclic-symmetric; and so is the function $h(x, y, z) = f(x, y, z) - f(y, z, x)$.

The two equations (*) defining a cyclic-symmetric function can be replaced by a single equation, if we use *place permutations* instead of *variable permutations*. Let τf be the function obtained from f by replacing the variable in the first entry of f by the variable in the third entry, the variable in the second entry by the variable in the first entry, and the variable in the third entry by the variable in the second

entry. Thus

$$\tau f(x, y, z) = f(z, x, y) \quad \text{and} \quad \tau f(x, z, y) = f(y, x, z).$$

Then we have the (immediately obvious)

PROPOSITION. *A function f is cyclic-symmetric if and only if $f + \tau f + \tau^2 f = 0$.*

III. Projections. We define three operators E_s, E_a, and E_c from the vector space V to itself, as follows. Set

$$E_s f = \frac{1}{6}\sum_{\sigma} \sigma f; \qquad E_a f = \frac{1}{6}\sum_{\sigma} (\operatorname{sign} \sigma)\sigma f,$$

where the sum ranges over all permutations σ of the set $\{x, y, z\}$.

To define the third operator E_c, let $\sigma f(x, y, z) = f(z, x, y)$. Set

$$E_c f = f - \frac{1}{3}(f + \sigma f + \sigma^2 f).$$

These operators were first defined by Alfred Young. It is easily verified that they are mutually orthogonal idempotents and that $E_s + E_a + E_c = \sigma_0$, where σ_0 is the identity operator. Furthermore, the function $E_s f$ is symmetric and the function $E_a f$ is skew-symmetric.

What Young did not see, however, is the following:

PROPOSITION. *For any function of three variables f the function $E_c f$ is cyclic-symmetric. Furthermore, if f is a cyclic-symmetric function, then $E_c f = f$.*

Proof. As to the first assertion, let $g = E_c f$. In other words, let

$$g(x, y, z) = f(x, y, z) - \frac{1}{3}[f(x, y, z) + f(z, x, y) + f(y, z, x)].$$

To show that the function g is cyclic-symmetric we must verify the identities:

$$g(x, y, z) + g(z, x, y) + g(y, z, x) = 0$$
$$g(x, z, y) + g(y, x, z) + g(z, y, x) = 0.$$

Let us verify the first of these identities. The left side expands to

$$\begin{aligned} &f(x, y, z) - \frac{1}{3}[f(x, y, z) + f(z, x, y) + f(y, z, x)] \\ &\quad + f(z, x, y) - \frac{1}{3}[f(z, x, y) + f(y, z, x) + f(x, y, z)] \\ &\quad + f(y, z, x) - \frac{1}{3}[f(y, z, x) + f(x, y, z) + f(z, x, y)]. \end{aligned}$$

Add each column to obtain

$$\begin{aligned} &f(x, y, z) + f(z, x, y) + f(y, z, x) \\ &\quad - \frac{1}{3}[f(x, y, z) + f(z, x, y) + f(y, z, x)] \\ &\quad - \frac{1}{3}[f(z, x, y) + f(y, z, x) + f(x, y, z)] \\ &\quad - \frac{1}{3}[f(y, z, x) + f(x, y, z) + f(z, x, y)] = 0. \end{aligned}$$

REFERENCES

1. A. Young, Collected Papers, The University of Toronto Press, 1977.
2. D. E. Littlewood, A University Algebra, 2nd edition, Whitefriars Press, London, 1961.
3. _____, Theory of Group Characters, 2nd edition, University Press, Oxford, 1958.
4. D. E. Rutherford, Substitutional Analysis, University Press, Edinburgh, 1948.
5. G. D. James and A. Kerber, The Representation Theory of the Symmetric Group, Cambridge University Press, New York, 1981.
6. F. D. Grosshans, G.-C. Rota and J. Stein, Invariant Theory and Superalgebras, CBMS, Amer. Math. Soc., 1987.

Irrationality Without Number Theory

RICHARD BEIGEL*
Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218

1. Introduction. Let k and t be positive integers. It is well known that $k^{1/t}$ is either a positive integer or an irrational number. This can be proved easily from the unique factorization theorem.

An interesting question is how much number theory is necessary in order to prove this theorem. Maier and Niven [**1**] and Floyd [**5**] have presented proofs of this theorem that use no facts about prime numbers. The former use the division algorithm and induction to simplify and generalize Steinhaus's proof [**4**, pp. 38–39] of the case $k = t = 2$; the latter uses the Euclidean algorithm to simplify and generalize Sagher's proof [**3**] of the case $t = 2$. When $t = 2$, Maier and Niven's technique is especially interesting because it does not explicitly use any number theory, only very basic inequalities.

In this paper, we also prove the theorem for all k and t. Although our proof is more complicated than Floyd's, it has the advantage of not explicitly using any number theory, so it can be presented to a very general audience. Our proof is simpler than Maier and Niven's. To accustom the reader to the technique involved, we begin by presenting a very simple proof for the case $t = 2$. The idea is due to a proof presented by Niven [**2**] for the case $k = t = 2$.

PROPOSITION 1. *$k^{1/2}$ is either a positive integer or an irrational number.*

Proof. Let $x = k^{1/2}$ and assume that x is rational. Then choose the smallest positive integer n such that nx is an integer. Let

$$n' = n(x - \lfloor x \rfloor).$$

where $\lfloor x \rfloor$ denotes the integer part (floor) of x. Since $0 \leqslant x - \lfloor x \rfloor < 1$, it follows that $0 \leqslant n' < n$. Note that $n' = nx - n\lfloor x \rfloor$, which is an integer. What's more,

$$n'x = nx^2 - (nx)\lfloor x \rfloor,$$

*Supported in part by grants CCR-8808949 and CCR-8958528 from the National Science Foundation.

REFERENCES

1. A. Young, Collected Papers, The University of Toronto Press, 1977.
2. D. E. Littlewood, A University Algebra, 2nd edition, Whitefriars Press, London, 1961.
3. _____, Theory of Group Characters, 2nd edition, University Press, Oxford, 1958.
4. D. E. Rutherford, Substitutional Analysis, University Press, Edinburgh, 1948.
5. G. D. James and A. Kerber, The Representation Theory of the Symmetric Group, Cambridge University Press, New York, 1981.
6. F. D. Grosshans, G.-C. Rota and J. Stein, Invariant Theory and Superalgebras, CBMS, Amer. Math. Soc., 1987.

Irrationality Without Number Theory

RICHARD BEIGEL*
Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218

1. Introduction. Let k and t be positive integers. It is well known that $k^{1/t}$ is either a positive integer or an irrational number. This can be proved easily from the unique factorization theorem.

An interesting question is how much number theory is necessary in order to prove this theorem. Maier and Niven [**1**] and Floyd [**5**] have presented proofs of this theorem that use no facts about prime numbers. The former use the division algorithm and induction to simplify and generalize Steinhaus's proof [**4**, pp. 38–39] of the case $k = t = 2$; the latter uses the Euclidean algorithm to simplify and generalize Sagher's proof [**3**] of the case $t = 2$. When $t = 2$, Maier and Niven's technique is especially interesting because it does not explicitly use any number theory, only very basic inequalities.

In this paper, we also prove the theorem for all k and t. Although our proof is more complicated than Floyd's, it has the advantage of not explicitly using any number theory, so it can be presented to a very general audience. Our proof is simpler than Maier and Niven's. To accustom the reader to the technique involved, we begin by presenting a very simple proof for the case $t = 2$. The idea is due to a proof presented by Niven [**2**] for the case $k = t = 2$.

PROPOSITION 1. *$k^{1/2}$ is either a positive integer or an irrational number.*

Proof. Let $x = k^{1/2}$ and assume that x is rational. Then choose the smallest positive integer n such that nx is an integer. Let

$$n' = n(x - \lfloor x \rfloor).$$

where $\lfloor x \rfloor$ denotes the integer part (floor) of x. Since $0 \leqslant x - \lfloor x \rfloor < 1$, it follows that $0 \leqslant n' < n$. Note that $n' = nx - n\lfloor x \rfloor$, which is an integer. What's more,

$$n'x = nx^2 - (nx)\lfloor x \rfloor,$$

*Supported in part by grants CCR-8808949 and CCR-8958528 from the National Science Foundation.

which is an integer. Because n is the smallest positive integer with that property, n' must be 0, so $x - \lfloor x \rfloor = 0$. In other words x is an integer. ■

2. Main Results. THEOREM 2. *$k^{1/t}$ is either an integer or an irrational number.*

Proof. Let $x = k^{1/t}$, and assume that x is rational. Then choose the smallest positive integer n such that nx is an integer. For $0 \leqslant i \leqslant t - 1$, we make the following claim, which we prove by induction on i:

INDUCTIVE HYPOTHESIS. *$n^{t-i-1}x^{t-i}$ is an integer.*

The base case ($i = 0$) is obvious. The theorem follows from the final case ($i = t - 1$). It remains to establish the inductive step. Assume that the inductive hypothesis is true for $i - 1$ so that

$$n^{t-i}x^{t-i+1} \text{ is an integer.}$$

Let

$$z = n^{t-i-1}x^{t-i}.$$

We complete the induction by showing that z is an integer. Let

$$n' = n(z - \lfloor z \rfloor).$$

Then $0 \leqslant n' < n$. We note that n' is the difference of two integers because $nz = (nx)^{t-i}$, which is an integer by the initial assumption, and $n\lfloor z \rfloor$ is the product of two integers. Therefore n' is a nonnegative integer. What's more,

$$\begin{aligned} n'x &= nzx - n\lfloor z \rfloor x \\ &= n^{t-i}x^{t-i+1} - (nx)\lfloor z \rfloor, \end{aligned}$$

which is the difference of two integers (by the inductive hypothesis and the initial assumption), hence an integer. Since n is the smallest positive integer with that property, n' must be 0. Therefore $z - \lfloor z \rfloor$ is 0, so z is an integer, completing the induction. ■

Using a trick from [**1**], we can extend this proof directly to apply to all algebraic integers.

COROLLARY 3. *If x is an algebraic integer, then x is either a rational integer or an irrational number.*

Proof. Recall that an algebraic integer is by definition a zero of a monic polynomial with integer coefficients, i.e.,

$$x^t + \sum_{0 \leqslant i \leqslant t-1} c_i x^i = 0.$$

Assume that x is rational, and choose the smallest positive integer n such that nx is an integer. Then, as observed by Maier and Niven

$$\begin{aligned} n^{t-1}x^t &= -\sum_{0 \leqslant i \leqslant t-1} c_i n^{t-1}x^i \\ &= -\sum_{0 \leqslant i \leqslant t-1} c_i n^{t-1-i}(nx)^i, \end{aligned}$$

which is an integer because nx is an integer. This establishes the base case for the

induction in the preceding theorem. The remainder of the proof is exactly as before, so we conclude that x is an integer. ■

3. Generalizations. Floyd also shows that $k^{1/3}$ cannot be a root of a quadratic equation with rational coefficients unless $k^{1/3}$ is an integer. Using a different technique (but no number theory) we extend this to $k^{1/t}$ for all odd t. (The result is false for even t because $(2^{t/2})^{1/t}$ is a root of a quadratic equation with rational coefficients.)

COROLLARY 4. *Let t be odd. If $k^{1/t}$ is a root of a quadratic equation with rational coefficients then $k^{1/t}$ is an integer.*

Proof. Let $x = k^{1/t}$ and assume that x is a root of a quadratic equation with rational coefficients. By the quadratic formula, there exists a rational number p and a nonnegative rational number q such that $x = p \pm q^{1/2}$. If $q = 0$, then x is rational, and hence an integer by the preceding theorem, so we need only consider the case $q > 0$.

$$\begin{aligned} k &= (p \pm q^{1/2})^t \\ &= \sum_i \binom{t}{i} p^{t-i} q^{i/2} (\pm 1)^i \quad \text{by the binomial theorem} \\ &= \pm \sum_{i \text{ odd}} \binom{t}{i} p^{t-i} q^{\lfloor i/2 \rfloor} q^{1/2} + \sum_{i \text{ even}} \binom{t}{i} p^{t-i} q^{i/2} \\ &= \pm s q^{1/2} + r, \end{aligned}$$

where

$$s = \sum_{i \text{ odd}} \binom{t}{i} p^{t-i} q^{\lfloor i/2 \rfloor},$$

and

$$r = \sum_{i \text{ even}} \binom{t}{i} p^{t-i} q^{i/2}.$$

Therefore, $q^{1/2}$ is equal to the rational number $\pm(k - r)/s$, unless $s = 0$. Each term of s is nonnegative because p is raised to an even power and q is positive. Furthermore, the last term ($i = t$) of s is equal to $q^{\lfloor t/2 \rfloor}$, which is strictly positive. Therefore s is positive, so $q^{1/2}$ is rational, so x is rational. By the preceding theorem, x is an integer. ■

It is natural to ask whether $k^{1/t}$ can be a non-integral zero of an mth degree polynomial with rational coefficients. (If t has a divisor d such that $2 \leq d \leq m$, then the answer is yes, because $(2^{t/d})^{1/t}$ is a zero of $x^{m-d}(x^d - 2)$.) The preceding corollary resolves the special case $m = 2$. It is unclear whether the proof technique works when $m = 3$ or $m = 4$; however it is clear that the proof technique cannot be applied when $m \geqslant 5$, because there is no closed-form solution of the quintic equation. Because the statement of the general case requires some number theory, it seems likely that a resolution of the general case will require some number theory. We present an approach that works when $t = 5$ and $m = 3$.

COROLLARY 5. *If $k^{1/5}$ is a root of a cubic equation with rational coefficients then $k^{1/5}$ is an integer.*

Proof. Let $x = k^{1/5}$ and assume that x is a root of a cubic equation with rational coefficients. Then there exist rational numbers p, q, and r such that $x^3 + px^2 + qx + r = 0$. Identically,

$$x^5 - k = (x^3 + px^2 + qx + r)(x^2 - px + (p^2 - q)) + (2pq - p^3 - r)x^2 + (pr + q^2 - p^2q)x + qr - p^2r - k.$$

Since $x^5 - k = 0$ and $x^3 + px^2 + qx + r = 0$,

$$(2pq - p^3 - r)x^2 + (pr + q^2 - p^2q)x + qr - p^2r - k = 0.$$

Thus x satisfies a quadratic equation with rational coefficients, and we are done by the preceding corollary, unless each coefficient of the quadratic equation is zero. From the first coefficient, we find $r = 2pq - p^3$. Substituting into the second coefficient, we find $q^2 + p^2q - p^4 = 0$. By the quadratic formula, $q = \frac{1}{2}(-1 \pm 5^{1/2})p^2$. Therefore, $5^{1/2} = \pm(2q/p^2 + 1)$, which is rational (and hence an integer by the preceding proposition), unless $p = 0$. But if $p = 0$, then $q = 0$ and $r = 0$, so $x^3 = 0$. Thus x is an integer. ■

4. Open Problems. Prove with a minimum amount of number theory:

- If $k^{1/t}$ is a zero of an irreducible cubic polynomial with integer coefficients and t is not divisible by 3 then $k^{1/t}$ is an integer.
- If $k^{1/t}$ is a zero of an irreducible mth degree polynomial with integer coefficients and t is relatively prime to m then $k^{1/t}$ is an integer.

5. Discussion. In our proofs, we used induction and the floor operation $\lfloor x \rfloor$. Since x is assumed to be rational, that operation implicitly uses the division algorithm. Thus, although the presentation of our proofs is simple, the depth of mathematics implicitly used is the same as in [1].

In Maier and Niven's proofs, they used induction, the division algorithm, and the floor operation. Although they appear not to use any number theory in the case $t = 2$, they use the floor operation in that case. In fact, it would be very surprising if any proof of irrationality really used no number theory.

In Floyd's proofs, he used the Euclidean algorithm, which can be derived by using induction and the division algorithm. Thus, Floyd's proofs also use the same depth of mathematics as in our proofs and in [1].

6. Acknowledgements. This work stems from discussions with Dan Ullman and Robert Floyd. I am grateful to the latter for his hospitality while I was visiting Stanford University and for proofreading, to Jim Owings for proofreading, to Bill Gasarch for tracking down a copy of [1], to Dan Spielman for assistance with the open problems, and to Ivan Niven for his encouragement.

REFERENCES

1. E. A. Maier and I. Niven, A method of establishing certain irrationalities, *Mathematics Magazine*, 37 (1964) 208–210.
2. I. Niven, Surprising results in elementary mathematics—II; presented at the Meeting of the AMS and the MAA, New Orleans, January, 1986.
3. Y. Sagher, What Pythagoras could have done, *American Mathematical Monthly*, 95 (1988) 117.
4. H. Steinhaus, Mathematical Shapshots, Stechert-Hafner, New York, 3rd edition, 1969.
5. R. W Floyd, What else Pythagoras could have done, *American Mathematical Monthly*, 96 (1989) 67.

Another Elementary Approach to the Jordan Form

J. I. HALL
Department of Mathematics, Michigan State University, East Lansing, MI 48824

1. Introduction. Several elementary treatments of the Jordan form are available. The recent article by Väliaho [**3**] in this MONTHLY contains a proof and a good discussion of many other elementary proofs. The approach in this note differs from most of these earlier proofs in two ways. No restriction is made upon the underlying field, and the proof avoids complicated bookkeeping and awkward calculational verifications of linear independence. That which is new in the proof given here resulted from an effort to add these two features to the excellent account (following Filippov [**1**]) to be found in the book of Ortega [**2**]. The present treatment makes reasonably clear the fact that having a Jordan form is a natural property of a linear transformation. A linear transformation can be diagonalized when its eigenvector structure is very strong; there must be a basis composed of eigenvectors. As we see here, a linear transformation has a Jordan form (a weaker property) when its eigenvector structure may not be as strong but is still sufficiently rich.

2. Existence of the Jordan form for linear transformations. Let X be a vector space over the field K, and let $\mathbf{T}$: $X \to X$ be a K-linear transformation of X. The null space of $\mathbf{T}$, null($\mathbf{T}$), is the subspace of all vectors of X taken to $\underline{0}$ by $\mathbf{T}$. The range of $\mathbf{T}$ on X, range($\mathbf{T}$), is the subspace of all vectors of the form $\mathbf{T}(\underline{x})$, for some $\underline{x}$ in X. Of fundamental importance is the fact that the dimension of X over K is equal to the sum of the dimensions of null($\mathbf{T}$) and range($\mathbf{T}$). (See [**2**, 2.3.3, p. 62].) We denote by $\dim(U)$ the dimension of the K-subspace U of X.

A Jordan chain for $\mathbf{T}$ in X of length m associated with λ is a set of non-zero vectors $\underline{x}_1, \ldots, \underline{x}_m$ such that

$$\mathbf{T}(\underline{x}_i) = \lambda \underline{x}_i + \underline{x}_{i-1}, \quad \text{for } i = 1, \ldots, m,$$

where we set $\underline{x}_0 = \underline{0}$. Note that if m is at least 1, then λ is an eigenvalue for $\mathbf{T}$ with associated eigenvector $\underline{x}_1$. In particular, an eigenvector by itself is a Jordan chain of length 1. If $\lambda = 0$, then $\mathbf{T}(\underline{x}_i) = \underline{x}_{i-1}$ for $i = 1, \ldots, m$; and the chain is nothing other than a "nil-orbit" for $\mathbf{T}$ in X. A Jordan chain for $\mathbf{T}$ associated with λ is also a chain associated with $\lambda + \alpha$ for $\mathbf{T} + \alpha\mathbf{I}$, where $\mathbf{I}$ is the identity linear transformation. In particular a chain for $\mathbf{T}$ associated with λ is merely a nil-orbit for $\mathbf{T} - \lambda\mathbf{I}$.

THEOREM 1. (*Existence of the Jordan form for linear transformations*). *Let* $\mathbf{A}$: $V \to V$ *be a linear transformation of the finite-dimensional* K*-vector space* V. *Assume that every non-zero* $\mathbf{A}$*-invariant subspace of* V *contains an eigenvector for* $\mathbf{A}$. *Then* V *possesses a basis which is a disjoint union of Jordan chains for* $\mathbf{A}$.

The proof is based upon a general proposition and a technical lemma about linear transformations which possess a Jordan form.

PROPOSITION 2. *Let* $\mathbf{T}$: $X \to X$ *be a linear transformation of the finite-dimensional* K*-vector space* X. *Let* W *be a* $\mathbf{T}$*-invariant subspace of* X, *and let* $\mathbf{S}$: $W \to W$ *be the restriction of* $\mathbf{T}$ *to* W. *Assume that*:

(i) W *contains the null space of* $\mathbf{T}$;

(ii) *The range of* $\mathbf{S}$ *on* W *equals the range of* $\mathbf{T}$ *on* X.

Then W *equals* X.

Proof. Indeed:

$$\begin{aligned}\dim(X) &= \dim(\text{null}(\mathbf{T})) + \dim(\text{range}(\mathbf{T})) \\ &= \dim(\text{null}(\mathbf{S})) + \dim(\text{range}(\mathbf{S})) = \dim(W).\end{aligned}$$

LEMMA 3. *Let* $\mathbf{T}$: $X \to X$ *be a linear transformation of the finite-dimensional* K*-vector space* X. *Assume that* X *has a basis which is a disjoint union of Jordan chains for* $\mathbf{T}$:

$$\begin{matrix} \underline{x}_{1,1}, \cdots, \underline{x}_{1,m_1} \\ \underline{x}_{2,1}, \cdots, \underline{x}_{2,m_2} \\ \vdots \\ \underline{x}_{k,1}, \cdots, \underline{x}_{k,m_k} \end{matrix};$$

where the i *th chain* $\underline{x}_{i,1}, \ldots, \underline{x}_{i,m_i}$ *has length* m_i *and is associated with the eigenvalue* λ_i. *Assume further that* $\lambda_1 = \cdots = \lambda_d = 0$ *and* $\lambda_i \neq 0$ *for* $i > d$. *Then*

(1) *null*$(\mathbf{T})$ *has as a basis* $\underline{x}_{1,1}, \ldots, \underline{x}_{i,1}, \ldots, \underline{x}_{d,1}$ *for* $1 \leqslant i \leqslant d$;

(2) *range*$(\mathbf{T})$ *has as basis the disjoint union of the Jordan chains*:

$$\begin{matrix} \underline{x}_{1,1}, \cdots, \underline{x}_{1,m_1-1} \\ \vdots \\ \underline{x}_{d,1}, \cdots, \underline{x}_{d,m_d-1} \\ \underline{x}_{d+1,1}, \cdots, \underline{x}_{d+1,m_{d+1}} \\ \vdots \\ \underline{x}_{k,1}, \cdots, \underline{x}_{k,m_k}. \end{matrix}$$

Proof. Clearly the given linearly independent sets are contained in these subspaces as claimed. Thus $\dim(\text{null}(\mathbf{T})) \geqslant d$ and $\dim(\text{range}(\mathbf{T})) \geqslant \dim(X) - d$. On the other hand

$$\dim(\text{null}(\mathbf{T})) + \dim(\text{range}(\mathbf{T})) = \dim(X);$$

so in each case we must have equality. The sets are bases.

Proof of Theorem 1. The proof is by induction on the dimension of the underlying space V. If $\mathbf{A}$ is a multiple of $\mathbf{I}$ (in particular when $\dim(V) = 1$), any basis is a basis of Jordan chains of length 1. Assume that this is not the case and that the theorem is true in all dimensions less than $n = \dim(V)$.

By assumption $\mathbf{A}$ has an eigenvector $\underline{v} \neq \underline{0}$ associated with some eigenvalue λ. Let $\mathbf{B}$ be the non-zero linear transformation $\mathbf{A} - \lambda\mathbf{I}$. Also let $\mathbf{C}$: $U \to U$ be the linear transformation of $U = \text{range}(\mathbf{B})$ which is the restriction of $\mathbf{B}$ to U. Note that $\underline{v} \in \text{null}(\mathbf{B})$, so U has a dimension m strictly smaller than n. A $\mathbf{C}$-invariant subspace of U is invariant under both $\mathbf{B}$ and $\mathbf{A}$. Therefore by induction we may conclude that U has a basis $\mathscr{U}$ which is a disjoint union of Jordan chains for $\mathbf{C}$ (and so for $\mathbf{B}$ and $\mathbf{A}$ as well). Let the chains of $\mathscr{U}$ be $\mathscr{U}_i$ for $1 \leqslant i \leqslant k$. Here the ith chain $\mathscr{U}_i$, consisting of $\underline{v}_{i,1}, \ldots, \underline{v}_{i,m_i}$, is associated with the eigenvalue λ_i. Assume that $\lambda_1 = \cdots = \lambda_d = 0$ and $\lambda_i \neq 0$ for $i > d$.

For each i with $1 \leqslant i \leqslant d$, $\underline{v}_{i,m_i}$ is in $U = \text{range}(\mathbf{B})$; so we may choose a vector $\underline{v}_{i,m_i+1}$ such that $\mathbf{B}(\underline{v}_{i,m_i+1}) = \underline{v}_{i,m_i}$. Next we choose vectors $\underline{z}_j$ (for $1 \leqslant j \leqslant q$) such that

$$\underline{z}_1, \ldots, \underline{z}_q, \underline{v}_{1,1}, \ldots, \underline{v}_{d,1}$$

is a basis for null$(\mathbf{B})$ which has dimension $q + d = n - m$.

Consider now the set $\mathscr{V}$ which is the union of $\mathscr{U}$, all the $\underline{v}_{i,m_i+1}$ for $1 \leqslant i \leqslant d$, and all the $\underline{z}_j$ for $1 \leqslant j \leqslant q$. This set contains

$$m + d + q = m + d + (n - m - d) = n$$

vectors. We first note that $\mathscr{V}$ is a union of $k + q$ disjoint Jordan chains for **B**, being the original chains $\mathscr{U}_i$ (extended by $\underline{v}_{i,m_i+1}$ when $1 \leqslant i \leqslant d$) and the new null vectors $\underline{z}_j$ (chains of length 1).

Our construction of the candidate basis $\mathscr{V}$ is essentially that of [**1**, **2**]. To complete the theorem it remains only to prove that $\mathscr{V}$ is indeed a basis for V. Let W be the subspace of V spanned by $\mathscr{V}$. As W contains all the $\underline{z}_j$ as well as $\underline{v}_{1,1}, \ldots, \underline{v}_{d,1}$, W contains *null*(**B**). As **W** contains the basis $\mathscr{U}$ of U, the range of **B**, it is certainly **B**-invariant. Indeed because, in addition to $\mathscr{U}$, **W** also contains each $\underline{v}_{i,m_i+1}$, for $1 \leqslant i \leqslant d$, Lemma 3 shows that the restriction of **B** to W still has U as its range. By Proposition 2, W is in fact equal to V. Thus the set $\mathscr{V}$ of size n spans the space V of dimension n and so is a basis of V, as required. This completes the proof of Theorem 1.

3. Existence of the Jordan form for matrices. If λ is a member of the field K and m is a positive integer, then the matrix $J_m(\lambda)$ is the $m \times m$ matrix over K with λ's on the diagonal, 1's on the superdiagonal, and 0's elsewhere. A matrix with entries from K is said to be in *Jordan form* if it is a block diagonal matrix, with all diagonal blocks equal to $J_m(\lambda)$ for various m and λ.

THEOREM 4. (*Existence of the Jordan form for matrices*). *Let A be a matrix with entries from the field K. Assume that the characteristic polynomial of A factors into linear factors over K. Then there is an invertible matrix P with entries from K such that $P^{-1}AP$ is in Jordan form.*

Let $V = K^n$, the n-dimensional K-space of column vectors. Then naturally associated with the matrix A we have the linear transformation **A**: $V \to V$ given by $\mathbf{A}(\underline{v}) = A\underline{v}$. The present theorem for the matrix A will be proven by applying the previous theorem to the linear transformation **A**.

LEMMA 5. *Let A be an $n \times n$ matrix with entries from the field K. Assume that the characteristic polynomial of A factors into linear factors over K. Then any non-zero* **A***-invariant subspace W of $V = K^n$ contains an eigenvector.*

Proof. Let $\underline{w}_1, \ldots, \underline{w}_m, \underline{x}_{m+1}, \ldots, \underline{x}_n$ be a basis for $V = K^n$ containing the basis $\underline{w}_1, \ldots, \underline{w}_m$ for W. Define Q to be the matrix whose ith column is $\underline{w}_i$ when $1 \leqslant i \leqslant m$ and is $\underline{x}_i$ when $m + 1 \leqslant i \leqslant n$. Then $B = Q^{-1}AQ$ is block upper triangular with an $n - m \times m$ block of 0's in the lower left. The characteristic polynomial of B is the product of that of the two diagonal blocks and is equal to that of A. In particular, the upper left $m \times m$ block has a characteristic polynomial which splits into linear factors over K. Therefore B has an eigenvector in V of the form $\underline{v} = (v_1, \ldots, v_m, 0, \ldots, 0)^T$. Now

$$\underline{w} = Q\underline{v} = v_1\underline{w}_1 + \cdots + v_m\underline{w}_m$$

is an eigenvector of A belonging to W.

Proof of Theorem 4. By Theorem 1 and Lemma 5, V has a basis $\mathscr{V}$ which is a disjoint union of Jordan chains $\underline{v}_{i,1}, \ldots, \underline{v}_{i,m_i}$, for $1 \leqslant i \leqslant k$. Let P be the matrix

form over the complex numbers is all that is needed, the use of Proposition 2 may simplify calculations. As we have seen here, that is the case with the proofs of [**1**, **2**]. The existence proof of [**3**] can also be shortened by use of Proposition 2.

REFERENCES

1. A. F. Filippov, A short proof of the theorem on the reduction of a matrix to Jordan form, *Moscow Univ. Math. Bull.*, 26 (1971) 70–71.
2. James M. Ortega, Matrix Theory, Plenum Press, New York, 1987.
3. H. Väliaho, An elementary approach to the Jordan form of a matrix, this MONTHLY, 93 (1986) 711–714.

Borromean Circles Are Impossible

BERNT LINDSTRÖM
Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm, Sweden

HANS-OLOV ZETTERSTRÖM
National Defence Research Establishment, S-171290 Sundbyberg, Sweden

The arrangement of three topological circles in R^3 depicted in FIG. 1 is called the *Borromean links* in knot theory [cf. **1**]. Such an arrangement of 'circles' is possible with curves, which are homeomorphic to geometric circles. To our knowledge it has not been observed before that the arrangement is impossible in R^3 if the circles are geometric even when the radii are arbitrary.

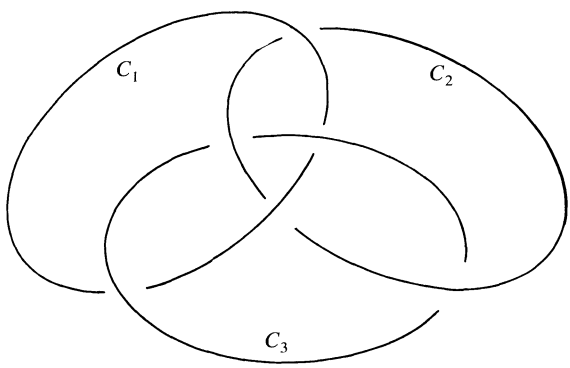


FIG. 1.

Our proof is indirect. We assume that the arrangement is possible using ordinary circles. Lift C_2 (cf. FIG. 1) towards the spectator such that it will meet C_1 in two points in its new position (C_2'). Observe that C_2' will not meet C_3 since we lift C_2 away from C_3.

There are now two cases: either C_1 and C_2' belong to a plane π or they do not belong to the same plane. In the first case C_3 will be partly above the plane, partly

form over the complex numbers is all that is needed, the use of Proposition 2 may simplify calculations. As we have seen here, that is the case with the proofs of [1, 2]. The existence proof of [3] can also be shortened by use of Proposition 2.

REFERENCES

1. A. F. Filippov, A short proof of the theorem on the reduction of a matrix to Jordan form, *Moscow Univ. Math. Bull.*, 26 (1971) 70–71.
2. James M. Ortega, Matrix Theory, Plenum Press, New York, 1987.
3. H. Väliaho, An elementary approach to the Jordan form of a matrix, this MONTHLY, 93 (1986) 711–714.

Borromean Circles Are Impossible

BERNT LINDSTRÖM
Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm, Sweden

HANS-OLOV ZETTERSTRÖM
National Defence Research Establishment, S-171290 Sundbyberg, Sweden

The arrangement of three topological circles in R^3 depicted in FIG. 1 is called the *Borromean links* in knot theory [cf. 1]. Such an arrangement of 'circles' is possible with curves, which are homeomorphic to geometric circles. To our knowledge it has not been observed before that the arrangement is impossible in R^3 if the circles are geometric even when the radii are arbitrary.

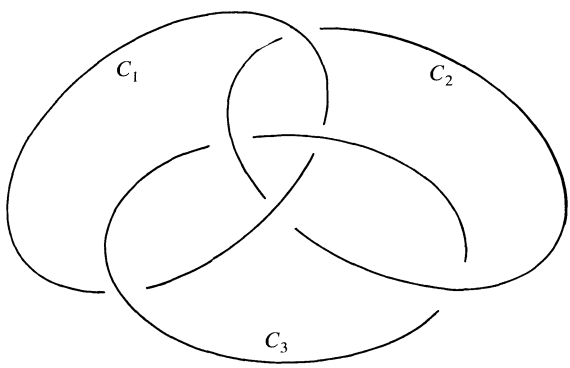


FIG. 1.

Our proof is indirect. We assume that the arrangement is possible using ordinary circles. Lift C_2 (cf. FIG. 1) towards the spectator such that it will meet C_1 in two points in its new position (C_2'). Observe that C_2' will not meet C_3 since we lift C_2 away from C_3.

There are now two cases: either C_1 and C_2' belong to a plane π or they do not belong to the same plane. In the first case C_3 will be partly above the plane, partly

below. If we follow the circle C_3 in clockwise direction we are first above the plane, then below, then above, then below, then back again and above the plane. This means that the circle C_3 meets the plane at least 4 times, which is impossible since the circle does not belong to the plane.

In the other case, when C_1 and C_2' do not lie in the same plane, we will prove that they belong to a sphere. The circle C_3 will meet this sphere at least 4 times, again a contradiction. We will now prove that C_1 and C_2' belong to a sphere when they are not coplanar. Let P and Q be the points in which C_1 and C_2' meet (see FIG. 2). Let R be the midpoint of the line segment PQ. Let S_1 and S_2 be the centres of the circles C_1 and C_2', respectively. The plane π' containing R, S_1 and S_2 is orthogonal to the line PQ. Let l_i be the normal to the plane of C_i through S_i ($i = 1, 2$). The line l_i is orthogonal to the line PQ and contains the point S_i in the plane π'. It follows that l_i lies in π' ($i = 1, 2$). Since l_1 and l_2 are not parallel lines they will meet in a point T. The distance between T and any point in $C_1 \cup C_2'$ is a constant r. Therefore C_1 and C_2' belong to the sphere with centre T and radius r. It is not hard to see that the circle C_3 goes in and out of this sphere 4 times, which is impossible because a circle and a sphere have at most two points in common when the circle does not belong to the sphere. Therefore, Borromean *circles* are impossible!

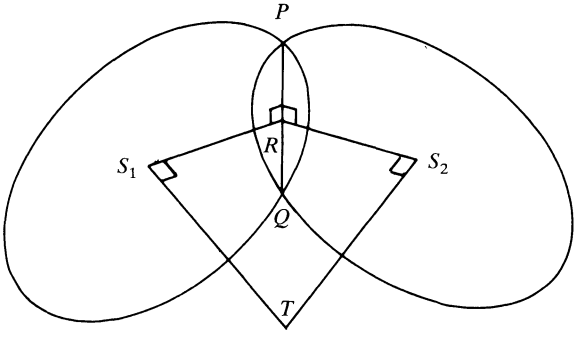


FIG. 2.

It is interesting to make a model of the Borromean rings in some elastic material like thin iron wire. Try to make the 'circles' as perfect as possible! The arrangement of rings will then attain infinitely many stable positions in space.

REFERENCE

1. D. Rolfsen, Knots and Links, Publish or Perish, Berkeley, 1976.

The Inclusion $L^p(\mu) \subseteq L^q(\nu)$

A. G. Miamee
Department of Mathematics, Hampton University, Hampton, VA 23668

Let Ω be a nonempty set and Σ be a σ-algebra of subsets of Ω. It is well known that $l^p(\Omega) \subseteq l^q(\Omega)$ whenever $0 < p \leqq q \leqq \infty$. Subramanian [3] characterized all positive measures μ on (Ω, Σ) for which $L^p(\mu) \subseteq L^q(\mu)$ whenever $0 < p \leqq q \leqq \infty$. On the other hand it is well known that if $\mu(\Omega) < \infty$ one has $L^q(\mu) \subseteq L^p(\mu)$ whenever $0 < p \leqq q \leqq \infty$. Romero [1] considered this second inclusion property and characterized all positive measures μ for which $L^q(\mu) \subseteq L^p(\mu)$ whenever $0 < p \leqq q \leqq \infty$.

In this note we consider the more general inclusion $L^p(\mu) \subseteq L^q(\nu)$, where now we have two measures involved and give a characterization for this general inclusion to hold (see the Main Theorem). Our proof for the Main Theorem is a nice application of the closed graph theorem.

Then we apply the Main Theorem to some interesting special cases and get more concrete results including some other versions of the results of Subramanian and Romero mentioned above. Our proofs for the results considered in [1] and [3] seem more to the point.

Concerning the definition of L^p spaces and the above inclusion the following remark is in order.

Remark. When $L^p(\mu)$ is regarded as a metric space; and that is what we would be doing throughout this work; the space which is really under consideration is not a space whose elements are functions, but a space whose elements are equivalence classes of functions. This is specially important here, and thus whenever we write $L^p(\mu) \subseteq L^q(\nu)$ we mean that every equivalence class of functions in $L^p(\mu)$ belongs to $L^q(\nu)$ as an equivalence class. There is, however, another possible interpretation for $L^p(\mu) \subseteq L^q(\nu)$, namely, any individual function f with $\int |f(\omega)|^p \, d\mu(\omega) < \infty$ has the property $\int |f(\omega)|^q \, d\nu(\omega) < \infty$.

The following lemma reveals the relation between these two different possible meanings of the inclusion $L^p(\mu) \subseteq L^q(\nu)$. Having this lemma we will interchange these two different meanings of $L^p(\mu) \subseteq L^q(\nu)$ in some appropriate settings, without mentioning the Lemma explicitly.

Lemma. *Let p and q be two positive real numbers. Then the inclusion $L^p(\mu) \subseteq L^q(\nu)$ holds in the sense of equivalence classes if and only if μ and ν are absolutely continuous with respect to each other and $L^p(\mu) \subseteq L^q(\nu)$ in the sense of individual functions.*

Proof. We just prove the "only if" part because the other part is easy. Suppose $L^p(\mu) \subseteq L^q(\nu)$ in its natural meaning as equivalence classes. Let E be a set in Σ with $\mu(E) = 0$ then χ_E is in the equivalence class 0 of $L^p(\mu)$ but this equivalence class 0 must belong to $L^q(\nu)$, by our hypothesis. Then 0 is necessarily the 0 equivalence class of $L^q(\nu)$. Now $\chi_E \in 0$ implies that $\chi_E = 0$ a.e. ν, which means $\nu(E) = 0$. Thus $\nu \ll \mu$. Similarly one can see $\mu \ll \nu$.

This work was supported by NASA Grant NAG-1-768 and by ONR Grant N00014-89-J-1824.

MAIN THEOREM. *Let p and q be any two positive real numbers. Then $L^p(\mu) \subseteq L^q(\nu)$ if and only if μ is absolutely continuous with respect to ν and there exists a constant $C(p,q)$ such that*

$$\|f\|_{q,\nu} \leqq C(p,q)\|f\|_{p,\mu}, \text{ for all } f \in L^p(\mu). \tag{1}$$

Proof. Since the "if" part is not hard to see we only show the "only if" part. Assume $L^p(\mu) \subseteq L^q(\nu)$. By the Lemma the two measures μ and ν are absolutely continuous with respect to each other. To show (1), first take the case that $p, q \geqq 1$, and define the linear operator I from the Banach space $L^p(\mu)$ into the Banach space $L^q(\nu)$ by

$$I(f) = f, \quad \text{for each } f \text{ in } L^p(\mu).$$

We claim that the operator I is closed (for the definition, see p. 120 [**2**]). To see this let f_n be a sequence in $L^p(\mu)$ such that

$$f_n \to f \text{ in } L^p(\mu)\text{-norm, and } I(f_n) \to g \text{ in } L^q(\nu)\text{-norm,}$$

for some f in $L^p(\mu)$ and some g in $L^q(\nu)$. One can then find a subsequence f_{n_i} of f_n such that (see for example Theorem 3.12 in [**2**])

$$f_{n_i} \to f \quad \text{a.e. } \mu, \text{ and } f_{n_i} \to g \quad \text{a.e. } \nu.$$

Now since ν is absolutely continuous with respect to μ we can write

$$f_{n_i} \to f \quad \text{a.e. } \nu, \text{ and } f_{n_i} \to g \quad \text{a.e. } \nu.$$

Thus $f = g$ a.e. ν, which means that f and g, regarded as elements of $L^q(\nu)$, are the same. This shows that the linear operator I is closed and hence bounded by the closed graph theorem (see for example p. 120 [**2**]). Thus there exists some constant $C(p,q)$ such that

$$\|f\|_{q,\nu} = \|I(f)\|_{q,\nu} \leqq C(p,q)\|f\|_{p,\mu}, \text{ for all } f \in L^p(\mu).$$

This completes the proof of (1) for the case $p, q \geqq 1$. To prove (1) for the other case let $r = \min(p,q)$, $p' = p/r$ and $q' = q/r$. One can easily show that the assumption $L^p(\mu) \subseteq L^q(\nu)$ implies $L^{p'}(\mu) \subseteq L^{q'}(\nu)$. But since clearly $p', q' \geqq 1$ we can use the result for the first case to write (note that $f \in L^p(\mu) \Rightarrow |f|^r \in L^{p'}(\mu)$)

$$\| |f|^r\|_{q',\nu} \leqq C(p',q')\| |f|^r\|_{p',\mu}, \text{ for all } f \in L^p(\mu)$$

for some constant $C(p',q')$. This can be written as

$$\|f\|_{q,\nu}^r \leqq C(p',q')\|f\|_{p,\mu}^r, \text{ for all } f \in L^p(\mu),$$

or as $\|f\|_{q,\nu} \leqq C(p,q)\|f\|_{p,\mu}$, for all $f \in L^p(\mu)$, where $C(p,q) = (C(p',q'))^{1/r}$.

Now we apply our Main Theorem to a couple of special cases to get more concrete results in these interesting cases.

PROPOSITION 1. *The following statements are equivalent*

(i) $L^p(\mu) \subseteq L^p(\nu)$ *for some positive real number* p.

(ii) μ *is absolutely continuous with respect to* ν *and there exists a constant* D *such that*

$$\nu(E) \leqq D\mu(E), \text{ for every set } E \text{ in } \Sigma. \tag{2}$$

(iii) $L^1(\mu) \subseteq L^1(\nu)$

(iv) $L^p(\mu) \subseteq L^p(\nu)$ *for every positive real number* p.

Proof. (i) $\Rightarrow$ (ii): The absolute continuity of μ with respect to ν follows from the Lemma. Since $L^p(\mu) \subseteq L^p(\nu)$, by the Main Theorem there exists a constant $C(p)$ such that

$$\|f\|_{p,\nu} \leqq C(p)\|f\|_{p,\mu}, \text{ for all } f \text{ in } L^p(\mu). \tag{3}$$

Take any set E in Σ with $\mu(E) < \infty$, then χ_E belongs to $L^p(\mu)$. Hence we can apply (3) to $f = \chi_E$ to get $\nu(E) \leqq D\mu(E)$ with $D = C(p)^p$. For the other E's the inequality (2) is obvious. (ii) $\Rightarrow$ (iii). Start with a simple function $f = \sum_{k=1}^{n} a_k \chi_{E_k}$ in $L^1(\mu)$, with E_k and E_j disjoint if $k \neq j$. Using (2) we can write

$$\|f\|_{1,\nu}^1 = \sum_{k=1}^{n} |a_k|\nu(E_k) \leqq D \sum_{k=1}^{n} |a_k|\mu(E_k) = D\|f\|_{1,\mu}^1, \tag{4}$$

which means that any simple function in $L^1(\mu)$ belongs to $L^1(\nu)$. Now take any function f in $L^1(\mu)$, and take a sequence f_n of simple functions in $L^1(\mu)$ such that $f_n \to f$ in $L^1(\mu)$-norm. Then f_n is clearly a Cauchy sequence in $L^1(\mu)$. Now (4) implies that f_n is also a Cauchy sequence in $L^1(\nu)$ and hence it converges to some function g in $L^1(\nu)$; i.e.,

$$f_n \to g \text{ in } L^1(\nu)\text{-norm}.$$

Using a subsequence argument similar to that in the proof of the Main Theorem one can show that $f = g$ a.e. ν. This together with the fact that g was in $L^1(\nu)$ implies that $f \in L^1(\nu)$. (iii) $\Rightarrow$ (iv): Let $f \in L^p(\mu)$. Then $|f|^p \in L^1(\mu)$. Thus assumption (iii) implies that $|f|^p \in L^1(\nu)$, which means that $f \in L^p(\nu)$. Finally, (iv) $\Rightarrow$ (i) is obvious.

Note. The proof shows that the constant D in part (ii) can be chosen to be independent of p, since we can take $D = C(1)$.

PROPOSITION 2. *The following statements are equivalent*

(i) $L^p(\mu) \subseteq L^q(\mu)$ *for some pair* p *and* q *of positive real numbers with* $p < q$.

(ii) *There exists a constant* $m > 0$ *such that*

$$\mu(E) \geqq m, \text{ for every } \mu\text{-non-null set } E \text{ in } \Sigma. \tag{5}$$

(iii) $L^p(\mu) \subseteq L^q(\mu)$ *for every pair* p *and* q *of positive real numbers with* $p \leqq q$.

Proof. (i) $\Rightarrow$ (ii): According to the Main Theorem there exists some positive constant C such that

$$\|f\|_{q,\mu} \leqq C\|f\|_{p,\mu}, \text{ for every } f \text{ in } L^p(\mu). \tag{6}$$

Let E be a μ-non-null set in Σ whose μ-measure is finite. Then $f = \chi_E$ belongs to $L^p(\mu)$. So by applying (6) to $f = \chi_E$ one gets $\mu(E)^{1/q} \leqq C\mu(E)^{1/p}$. Since $1/p > 1/q$ (because $p < q$) and $0 < \mu(E) < \infty$ one gets $\mu(E)^{1/p-1/q} \geqq 1/C$ which implies (5) with $m = C^{pq/(p-q)}$. (ii) $\Rightarrow$ (iii): Let $f \in L^p(\mu)$. For any positive integer n let $E_n = \{\omega \in \Omega: |f(\omega)| > n\}$. Since

$$n^p\mu(E_n) \leqq \int_{E_n} |f|^p \, d\mu \leqq \int_{\Omega} |f|^p \, d\mu < \infty,$$

we have $\mu(E_n) < \infty$ for every n. Thus by the hypothesis in (ii) either $\mu(E_n) = 0$ or $\mu(E_n) \geqq m$. On the other hand $E_{n+1} \subseteq E_n$, $\mu(E_1) < \infty$, and $\bigcap_{n=1}^{\infty} E_n = \varnothing$ imply that $\mu(E_n) \to 0$. Thus there must exist some positive integer n_0 such that $|f(\omega)| \leqq$

n_0 for μ-almost all $\omega \in \Omega$. Hence

$$\int_\Omega |f|^q \, d\mu = \int_\Omega |f|^p |f|^{q-p} \, d\mu \leqq n_0^{q-p} \int |f|^p \, d\mu < \infty.$$

(iii) $\Rightarrow$ (i): This is obvious.

PROPOSITION 3. *The following statements are equivalent.*

(i) $L^q(\mu) \subseteq L^p(\mu)$ *for some pair of positive real numbers* p *and* q *with* $p < q$.

(ii) *There exists some constant* $M > 0$ *such that*

$$\mu(E) \leqq M, \text{ for every set } E \text{ in } \Sigma \text{ with finite } \mu\text{-measure}. \tag{7}$$

(iii) $L^q(\mu) \subseteq L^p(\mu)$ *for every pair* p *and* q *of positive numbers with* $p \leqq q$.

Proof. (i) $\Rightarrow$ (ii): Using the Main Theorem, one can see (as in the proof of (i) $\Rightarrow$ (ii) in Proposition 2) that there exists a constant $C > 0$ such that

$$\mu(E)^{1/p} \leqq C\mu(E)^{1/q}, \text{ for every set } E \text{ in } \Sigma \text{ with finite } \mu\text{-measure}.$$

Since $1/p > 1/q$, if E is some μ-non-null set in Σ of finite μ-measure we can write $\mu(E)^{1/p-1/q} \leqq C$ which means (7) holds with $M = C^{pq/q-p}$. (ii) $\Rightarrow$ (iii): Let M be the smallest number for which (7) holds. There exists a sequence F_n of sets in Σ with finite measure such that $\mu(F_n) \to M$ as $n \to \infty$. Let $S = \bigcup_{n=1}^{\infty} F_n$ then $\mu(S) = \mu(\bigcup_{n=1}^{\infty} F_n) = \lim_{m\to\infty} \mu(\bigcup_{n=1}^{m} F_n)$. But by the choice of M, $\mu(\bigcup_{n=1}^{m} F_n) \leqq M$ for every m. So $\mu(S) \leqq M$. On the other hand $\mu(S) \geqq \mu(F_n)$ for every n and $\mu(F_n) \to M$. This gives $\mu(S) \geqq M$. Thus $\mu(S) = M$. Letting $T = \Omega - S$ one concludes that either $\mu(T) = 0$ or T is an atom of infinite measure. Now let f be in $L^p(\mu)$, then f must vanish on T almost everywhere μ. So

$$\int_\Omega |f|^p \, d\mu = \int_S |f|^p \, d\mu + \int_T |f|^p \, d\mu = \int_S |f|^p \, d\mu.$$

Since $\mu(S) = M < \infty$ we can apply Hölder's inequality to get

$$\int_\Omega |f|^p \, d\mu \leqq \mu(S)^{(q-p)/q} \left(\int_S |f|^q \, d\mu \right)^{p/q} < \infty.$$

That is to say $f \in L^p(\mu)$. Again (iii) $\Rightarrow$ (i) is clear.

We conclude this note by saying that most of these results and especially the Main Theorem is correct for the case of matrix-valued functions and the proof given here goes through in a similar fashion.

REFERENCES

1. J. I. Romero, When is $L^p(\mu)$ contained in $L^q(\mu)$?, *Amer. Math. Monthly*, 90 (1983) 203–206.
2. W. Rudin, Real and Complex Analysis, McGraw Hill, New York, 1974.
3. B. Subramanian, On the inclusion $L^p(\mu) \subseteq L^q(\mu)$, *Amer. Math. Monthly*, 85 (1978) 479–781.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

π and the Limit of $(\sin \alpha)/\alpha$

LEONARD GILLMAN
Department of Mathematics, University of Texas, Austin, TX 78712

1. Introduction. The formulas $C = 2\pi r$ and $A = \pi r^2$ are part of the vocabulary of our students and of many ordinary citizens, but very few know the ideas underlying them—not even my recent honors calculus class, who told me they were handed the formulas in high school without explanation. As for the presence of π in both formulas, some thoughtful students may consider it a miracle; but probably most people regard π, with its mysterious string of decimals, as just one more example of the depressing magic that math teachers dish out for them to memorize.

Devoting even one calculus or precalculus period to the fundamental ideas can accomplish several things: come to grips in a nonthreatening way with the completeness of the real numbers, afford a preview of the integral in a familiar context, clarify the basis for the trigonometric functions, show the theorem on the limit of $(\sin \alpha)/\alpha$ in its natural setting as essentially just the definition of the circumference of a circle, and of course elucidate the famous formulas themselves (which the students are going to remember long after they have forgotten all their calculus).

What follows is an expanded account of what I presented to the class (and later to an MAA section meeting). It was a pleasure to see their eyes light up at the punch line.

2. π and $(\sin \alpha)/\alpha$. The perimeter of any circumscribed polygon about a circle is greater than that of any inscribed polygon. (This is a good exercise for the students, though they may need hints.) Obviously, the area inside any circumscribed polygon exceeds that inside an inscribed polygon. It is convenient to restrict our attention to the inscribed regular polygons P_m, where m, the number of sides, is a power of 2. (I had my students draw the picture for $m = 4$, 8, 16, and 32.) We define C, the circumference of the circle, to be the least upper bound of the perimeters of the P_m, and A, the area inside the circle, to be the least upper bound of the areas they enclose.

Polygonal lengths on a circle of radius r are equal to r times the corresponding lengths for the unit circle. So, then, is their least upper bound. (Here and later we need the theorem that $\text{lub } cx = c \text{ lub } x$ $(c \geqslant 0)$; I would show students this proof.) The circumference of a circle is therefore proportional to the radius. The constant of proportionality—i.e., the circumference of the unit circle—is called 2π; this defines π. Thus, $C = 2\pi r$.

On the unit circle, a pair of consecutive vertices of P_m marks off an arc of length $2\pi/m$—at the present stage, this is an axiom rather than a theorem—and this number is defined to be the radian measure of the corresponding central angle. On a circle of radius r, the angle cuts off an arc of length $2\pi r/m$.

FIGURE 1 shows the sector associated with a side AB of a polygon P_m, and its two half-sectors corresponding to sides AD and DB of P_{2m}. The central half-angle

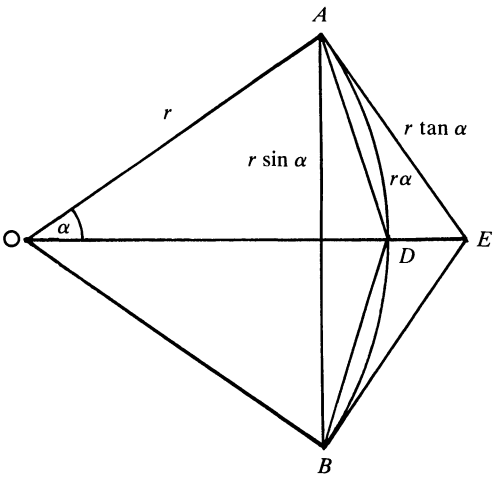


FIG. 1.

α (or α_{2m}) has arc $r\alpha$, and the half-side of P_m is $r \sin \alpha$. Since the ratio of side to arc is the same as perimeter to circumference,

$$\frac{\sin \alpha}{\alpha} = \frac{P_m}{C} \to 1. \tag{1}$$

The area inside triangle OAD is $(1/2)r^2 \sin \alpha$; and the area inside P_{2m} is $C/(r\alpha)$ times that area. Therefore,

$$\text{area } P_{2m} = \frac{1}{2} Cr \frac{\sin \alpha}{\alpha}.$$

Taking least upper bounds yields

$$A = \frac{1}{2} Cr. \tag{2}$$

(This is what so pleased the students.) With $C = 2\pi r$, $A = \pi r^2$.

3. Remarks on the proof. A variant derivation bypasses $(\sin \alpha)/\alpha$, using the more elementary fact that $\cos \alpha \to 1$ as $\alpha \to 0$. To verify this latter, assume for simplicity that $0 < \alpha < \pi/2$; then

$$0 < 1 - \cos \alpha < 1 - \cos^2 \alpha = \sin^2 \alpha < \sin \alpha < \alpha \tag{3}$$

—this last from FIGURE 1. By the squeeze theorem, $1 - \cos \alpha \to 0$, whence $\cos \alpha \to 1$, as $\alpha \to 0$. (Note too for later reference that $\sin \alpha \to 0$.)

In FIGURE 1, the area enclosed by triangle OAB is $(1/2)r^2 \sin 2\alpha$, and the length of AB, a side of P_m, is $2r \sin \alpha$. The area and perimeter of P_m are in the same ratio, which is $(r/2)\cos \alpha$. Taking limits, we get $A/C = r/2$. □

A heuristic form of these arguments is the well-known "cut-and-unroll" proof shown in FIGURE 2.

At the other extreme, [1] uses integration by parts to transform the (improper) integral for the length of arc of a quarter-circle to $(2/r)$ times the integral for the area.

upper bound of polygonal paths; the fact that congruent arcs then have the same length; the fact that arc length is additive (needed in deriving the formula for $\cos(\alpha + \beta)$); and the existence of arcs of arbitrary length, so that $\sin \alpha$ and $\cos \alpha$ really are defined for all real α [**2**, pp. 198–199]. Everyone will have one's own idea about whether, when, and how to bring them in. Most texts say nothing at all.

In our discussion, the length of arc on the unit circle determined by a side of P_m is defined—as C/m; therefore $\sin \alpha$ and $\cos \alpha$ are defined for $\alpha = 2\pi/4, 2\pi/8, 2\pi/16, \ldots$. Trivially, the sequence of perimeters $P_4, P_8, P_{16}, \ldots$ is increasing (which is why we stick to powers of 2); the limit notation in (1) is therefore intuitively clear. But is it reasonable to evaluate the limit of $(\sin \alpha)/\alpha$ (and the other limits) for α ranging merely over a sequence? Well, at least that's a more substantial domain than the empty set; moreover, we get the result free of charge in the course of deriving the formula $A = \pi r^2$. In any case, we are still permitted to show students the standard argument based on the areas of the right triangles and the sector.

By the way, if we now define the length along the unit circle determined by k consecutive sides of P_m to be kC/m—or, alternatively, if we decree additivity—then all dyadic rational multiples of C appear as arc lengths. This dense set of values, obtained effortlessly, gives the student something to hang onto. (One can then get the lengths of all other arcs from this set by means of least upper bounds.)

Acknowledgements. This article is an elaboration of material that was at one time part of a joint venture with R. H. McDowell, whom I wish to thank for permission to publish the material separately. I also wish to thank the referee for several helpful suggestions.

REFERENCES

1. E. F. Assmus, Jr., Pi, this MONTHLY, 92 (1985) 213–214.
2. W. F. Eberlein, The circular function(s), *Math. Mag.*, 39 (1966) 197–201.

The Structure of Orthogonal Transformations

LEE RUDOLPH[1]
Department of Mathematics, Clark University, Worcester, MA 01610

1. Introduction. What does an orthogonal transformation "look like" in Euclidean space of dimension n?

For $n \leqslant 2$, most of my linear algebra students can literally see the answer to this question once I tell them what an orthogonal transformation is—any linear transformation which leaves all lengths unchanged—and point out that they already are familiar with some geometrically defined examples: reflections and rotations.

In fact, an orthogonal transformation of a line ($n = 1$) is either the identity or the reflection across the origin, and an orthogonal transformation of a plane

[1]Partially supported by N.S.F. grant DMS-8801959.

upper bound of polygonal paths; the fact that congruent arcs then have the same length; the fact that arc length is additive (needed in deriving the formula for $\cos(\alpha + \beta)$); and the existence of arcs of arbitrary length, so that $\sin \alpha$ and $\cos \alpha$ really are defined for all real α [**2**, pp. 198–199]. Everyone will have one's own idea about whether, when, and how to bring them in. Most texts say nothing at all.

In our discussion, the length of arc on the unit circle determined by a side of P_m is defined—as C/m; therefore $\sin \alpha$ and $\cos \alpha$ are defined for $\alpha = 2\pi/4, 2\pi/8, 2\pi/16, \ldots$. Trivially, the sequence of perimeters $P_4, P_8, P_{16}, \ldots$ is increasing (which is why we stick to powers of 2); the limit notation in (1) is therefore intuitively clear. But is it reasonable to evaluate the limit of $(\sin \alpha)/\alpha$ (and the other limits) for α ranging merely over a sequence? Well, at least that's a more substantial domain than the empty set; moreover, we get the result free of charge in the course of deriving the formula $A = \pi r^2$. In any case, we are still permitted to show students the standard argument based on the areas of the right triangles and the sector.

By the way, if we now define the length along the unit circle determined by k consecutive sides of P_m to be kC/m—or, alternatively, if we decree additivity—then all dyadic rational multiples of C appear as arc lengths. This dense set of values, obtained effortlessly, gives the student something to hang onto. (One can then get the lengths of all other arcs from this set by means of least upper bounds.)

Acknowledgements. This article is an elaboration of material that was at one time part of a joint venture with R. H. McDowell, whom I wish to thank for permission to publish the material separately. I also wish to thank the referee for several helpful suggestions.

REFERENCES

1. E. F. Assmus, Jr., Pi, this MONTHLY, 92 (1985) 213–214.
2. W. F. Eberlein, The circular function(s), *Math. Mag.*, 39 (1966) 197–201.

The Structure of Orthogonal Transformations

LEE RUDOLPH[1]
Department of Mathematics, Clark University, Worcester, MA 01610

1. Introduction. What does an orthogonal transformation "look like" in Euclidean space of dimension n?

For $n \leqslant 2$, most of my linear algebra students can literally see the answer to this question once I tell them what an orthogonal transformation is—any linear transformation which leaves all lengths unchanged—and point out that they already are familiar with some geometrically defined examples: reflections and rotations.

In fact, an orthogonal transformation of a line ($n = 1$) is either the identity or the reflection across the origin, and an orthogonal transformation of a plane

[1]Partially supported by N.S.F. grant DMS-8801959.

($n = 2$) is either the reflection across some line or some rotation around the origin. Not only is it easy to prove these facts of low-dimensional Euclidean geometry, they are visually obvious to almost everyone.

The main point of this paper is to show how an orthogonal transformation in any dimension can (in a sense) be visualized using the known examples in the line and the plane. This is explained in §5, with the help of a theorem proved in §4: for $n > 0$, *every orthogonal transformation has an invariant line or plane.*

Of course these are familiar results; see, for instance, [**1**, §81]. However, the particular geometric proof given in §4 seems less well known (I would welcome a reference). Complex eigenvalues make no appearance. Instead, *Rayleigh quotients* play a central role; they are defined and differentiated in §3, after some basic material on adjoints is recalled in §2.

I have successfully presented this proof several times, in a linear algebra course peopled by math majors who have studied "vector calculus" but not yet "abstract algebra." My bias is always towards geometric methods; they seem especially appropriate to a study of the structure of orthogonal transformations. The visual intuition and physical experiments which are available in dimensions 1, 2, and 3 can be parlayed into a good understanding of the general case.

2. Adjoints. Let $\mathbb{E}_n$ be a Euclidean vectorspace of dimension n, with the inner product of vectors v and w denoted by $\langle v, w \rangle$. A linear transformation A of $\mathbb{E}_n$ is orthogonal if $\langle A(v), A(w) \rangle = \langle v, w \rangle$ for all v and w, and symmetric if $\langle A(v), w \rangle = \langle v, A(w) \rangle$ for all v and w.

Both definitions can be rephrased in terms of the *adjoint* of A, that is, the linear transformation of $\mathbb{E}_n$ determined uniquely by the requirement that $\langle A^*(v), w \rangle = \langle v, A(w) \rangle$ for all v and w: the transformation A is orthogonal if and only if $A^* = A^{-1}$, and symmetric if and only if $A^* = A$. (Symmetric transformations are, in fact, often called *self-adjoint*.)

Remark. For the convenience of the reader who is more at home with matrices, the definitions can be rephrased yet again. (Neither matrices nor bases will be used in the rest of the paper.) A basis $v_i (i = 1, \ldots, n)$ of $\mathbb{E}_n$ is orthonormal if $\langle v_i, v_j \rangle$ equals 1 or 0 according as i equals j or not. Let the (square) matrix of A with respect to some orthonormal basis be $[a_{ij}]$. Then the matrix of A^* with respect to the same basis is the transposed matrix $[a_{ji}]$. It follows that A is orthogonal if and only if the rows (or columns) of its matrix, with respect to any orthonormal basis of $\mathbb{E}_n$, are an orthonormal basis of the Euclidian vector space of row (or column) vectors with the usual "dot product" as inner product. Similarly, A is symmetric if and only if its matrix (with respect to any orthonormal basis) is its own transpose, that is, symmetric across its main diagonal.

3. Rayleigh quotients. The *Rayleigh quotient* of a non-zero vector v (with respect to A) is $R(v) := \langle A(v), v \rangle / \langle v, v \rangle$, that is, the scalar component of $A(v)$ along v.

LEMMA 1. *R attains maximum and minimum values.*

Proof. R is continuous, and attains all its values on the closed, bounded set of unit vectors (since $R(tv) = R(v)$ for $t \neq 0$). □

LEMMA 2. *R is differentiable, with gradient*

$$\nabla R(v) = \frac{A(v) + A^*(v) - 2R(v)v}{\langle v, v \rangle}.$$

Proof. It is sufficient (and necessary) to prove that

$$\langle \nabla R(v), w \rangle = \frac{\langle A(v) + A^*(v) - 2R(v)v, w \rangle}{\langle v, v \rangle}$$

for every vector w. The proof is straightforward calculation. It can be accomplished by using the quotient rule and twice applying (once with $B = A$, once with $B = I$) the identity

$$\langle \nabla \langle v, B(v) \rangle, w \rangle = \langle w, B(v) \rangle + \langle v, B(w) \rangle$$

which holds for any linear transformation B of $\mathbb{E}_n$. □

Remark. The Rayleigh quotient seems to be a very useful geometric tool. It is applied in the next section to the special cases of orthogonal and symmetric transformations. Another application is given in [2]: if A is an arbitrary linear transformation of $\mathbb{E}_n$, then the trace of A is n times the average value of its Rayleigh quotient on the hypersphere S^{n-1} of unit vectors.

4. Invariant subspaces of orthogonal transformations. A subspace $F \subset \mathbb{E}_n$ is *invariant* (for the linear transformation A) if $A(v) \in F$ for every $v \in F$. Of course $\mathbb{E}_n$ and $\{0\}$ are trivially invariant; any other invariant subspace is *non-trivial*. There are many linear transformations without non-trivial invariant subspaces.

Even an orthogonal transformation may not have any non-trivial invariant subspaces. For instance, $\mathbb{E}_1$ has no non-trivial subspaces to be invariant; more seriously, $\mathbb{E}_2$ has many non-trivial subspaces (all the lines through the origin), but none of them is invariant if A is rotation by an angle other than an integer multiple of π. However, those turn out to be the only problematic cases, as we see from the following theorem, promised in the introduction.

THEOREM 1. *If A is an orthogonal transformation of* $\mathbb{E}_n$ $(n > 0)$, *then A has at least one invariant subspace of dimension* 1 *or* 2 (*which is, of course, non-trivial if* $n > 2$).

Proof. Let v be a *critical point* of the Rayleigh quotient function of A, that is, $\nabla R(v) = 0$. (By Lemmas 1 and 2, we know that such points exist—for instance, the maxima and minima of R.) Then $A(v) + A^*(v) - 2R(v)v = 0$. This is a linear dependence among the three vectors $A(v)$, $A^*(v) = A^{-1}(v)$, and v, so the subspace they span has dimension 1 or 2, and is invariant. □

Remarks. (1) When A is orthogonal, $|R| \leqslant 1$. If $|R(v)| = 1$, then $A(v) = \pm v$, so v spans an invariant line. If the maximum value of $|R|$ is less than 1, then A has an invariant plane but no invariant line. (2) Suppose that, rather than orthogonal, A is symmetric. Then $A(v) + A^*(v) - 2R(v)v = 0$ is a linear dependence between the *two* vectors $A(v) = A^*(v)$ and v. We see that, if v is a critical point of R, then $A(v) = R(v)v$, so the effect of A is simply to multiply v by a scalar; conversely, if $A(v) = tv$ for some $v \neq 0$, then $R(v) = t$. In other words, when A is symmetric, the critical points of R are exactly the *eigenvectors* of A (and the critical values of R are its *eigenvalues*). This is the original, classic application of

Rayleigh quotients (see [3]) and might be a good geometric motivation for the study of eigenvectors fairly early in a course.

5. The structure of an orthogonal transformation. In this section we apply Theorem 1 to develop some geometric understanding of orthogonal transformations.

Let A be an orthogonal transformation of $\mathbb{E}_n$, $n > 0$. As a first step, we state the following obvious property of invariant subspaces for A.

LEMMA 3. *If F is an invariant subspace, then its orthogonal complement $F^\perp$ is also invariant, and $A|F$, the restriction of A to F, is an orthogonal transformation of F.*

The following terminology is convenient. Call the plane P a *rotation plane* (for A) if P is invariant and $A|P$ is some rotation about the origin, possibly the identity; call P a *reflection plane* if P is invariant and $A|P$ is reflection across some line. Every invariant plane is either a rotation plane or a reflection plane.

THEOREM 2. *If $n \geqslant 3$, then A has a rotation plane.*

Proof. If $n = 3$, then A has an invariant plane P (by Lemma 3 and Theorem 1). If P is not a rotation plane, then let $M \subset P$ be the line such that $A|P$ is a reflection across M. Either $A|P^\perp$ is the identity or it is a reflection across the origin. In the first case, the span Q of M and $P^\perp$ is a rotation plane ($A|Q$ is the identity); in the second case, $M^\perp$ is a rotation plane ($A|M^\perp$ is rotation by π).

If $n = 4$, then (by Theorem 1) either A has an invariant line L or A has an invariant plane P. In the first case, the Euclidean 3-space $L^\perp$ is invariant (by Lemma 3), and then, by the preceding paragraph, $A|L^\perp$ has a rotation plane. In the second case, either P is a rotation plane, or $P^\perp$ is a rotation plane, or both are reflection planes—but then the span of the reflection lines in P and $P^\perp$ is a rotation plane.

Finally, if $n \geqslant 5$, then A has an invariant subspace F of dimension at least 3 (by Theorem 1 and Lemma 3). By induction, $A|F$ has a rotation plane. □

COROLLARY. *If $n = 2k + 1$ (resp., $2k + 2$) is odd (resp., even), then $\mathbb{E}_n$ is the orthogonal direct sum of k rotation planes $P_1, \ldots, P_k$ and an invariant line L (resp., an invariant plane P_0).*

Thus, for example, in $\mathbb{E}_3$ every orthogonal transformation with positive determinant is a rotation around some line.

COROLLARY. *Any orthogonal transformation A of $\mathbb{E}_n$ can be expressed as a composition of n or fewer reflections in hyperplanes.*

Both corollaries are easily proved by induction.

Acknowledgements. I thank the referee for many helpful suggestions. Avner Ash introduced me to the phrase "Rayleigh quotient," and gave me the reference to [3], during a discussion of §4 between the first and second drafts of this paper.

REFERENCES

1. Paul Halmos, Finite-Dimensional Vector Spaces, Van Nostrand, 1958 (reissued by Springer).
2. Lee Rudolph, A trace of geometry, preprint (1989).
3. Gilbert Strang, Linear Algebra and Its Applications, 3rd ed., Harcourt Brace Jovanovich, 1988.

Using the property of comparability of conforming sets that we have just proved, we now observe that if A is a conforming subset of X and $x \in A$, then whenever $y < x$, either $y \in A$ or y does not belong to any conforming set. It now follows easily that the union U of all the conforming subsets of X is conforming, and we deduce from this fact that if $x = f(U)$, then the set $U \cup \{x\}$ is conforming. Therefore, $x \in U$, contradicting the fact that x is a strict upper bound of U.

REFERENCES

1. P. R. Halmos, Naive Set Theory, D. Van Nostrand Company, Inc., Princeton, NJ, 1960.
2. W. Rudin, Real and Complex Analysis, McGraw-Hill Book Company, New York, 1987.

The Converse of Liouville's Theorem

RICHARD GRANT
Departments of Geriatrics and Psychiatry, University of Massachusetts Medical School, Worcester, MA 01655

After seeing Liouville's theorem on the approximation of real algebraic numbers by rationals, it is natural to ask whether the converse is true. That is, if there are only finitely many rational numbers p/q such that $|\alpha - p/q| < c/q^k$ for some positive integer k and constant c, can we conclude that α is algebraic? There is a simple counterexample that isn't mentioned in the standard texts [**1**, **2**, **3**, **4**].

If $\mathbf{A} = \{[a_0, a_1, a_2, \ldots]; a_i = 1 \text{ or } 2\}$ is the set of numbers whose continued fraction representation contains only 1s and 2s then $\mathbf{A}$ is uncountable and so must contain transcendental numbers. However if $\alpha = [a_0, a_1, a_2, \ldots] \in \mathbf{A}$ then α has no close rational approximation. To see this let p_m/q_m be the mth convergent. For any simple continued fraction we have

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{1}{2}\left|\frac{p_{m+1}}{q_{m+1}} - \frac{p_m}{q_m}\right| = \frac{1}{2q_m q_{m+1}}.$$

But $q_{m+1} = q_m + q_{m-1}$ or $q_{m+1} = 2q_m + q_{m-1}$ since $a_m = 1$ or 2 so, for the numbers under consideration, we have $q_{m+1} \leqslant 3q_m$ and so

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{1}{6q_m^2}.$$

Moreover, this inequality holds for any rational number p/q, for if $|\alpha - p/q| \leqslant 1/6q^2$ then $|\alpha - p/q| < 1/2q^2$, and this implies that p/q is a convergent; see [**4**, p. 219].

REFERENCES

1. H. Davenport, The Higher Arithmetic, Dover, New York, 1983.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, London, 1960.
3. Hua Loo-Keng, Introduction to Number Theory, Springer, New York, 1982.
4. H. M. Stark, An Introduction to Number Theory, MIT Press, Cambridge, Massachusetts, 1981.

Using the property of comparability of conforming sets that we have just proved, we now observe that if A is a conforming subset of X and $x \in A$, then whenever $y < x$, either $y \in A$ or y does not belong to any conforming set. It now follows easily that the union U of all the conforming subsets of X is conforming, and we deduce from this fact that if $x = f(U)$, then the set $U \cup \{x\}$ is conforming. Therefore, $x \in U$, contradicting the fact that x is a strict upper bound of U.

REFERENCES

1. P. R. Halmos, Naive Set Theory, D. Van Nostrand Company, Inc., Princeton, NJ, 1960.
2. W. Rudin, Real and Complex Analysis, McGraw-Hill Book Company, New York, 1987.

The Converse of Liouville's Theorem

RICHARD GRANT
Departments of Geriatrics and Psychiatry, University of Massachusetts Medical School, Worcester, MA 01655

After seeing Liouville's theorem on the approximation of real algebraic numbers by rationals, it is natural to ask whether the converse is true. That is, if there are only finitely many rational numbers p/q such that $|\alpha - p/q| < c/q^k$ for some positive integer k and constant c, can we conclude that α is algebraic? There is a simple counterexample that isn't mentioned in the standard texts [**1**, **2**, **3**, **4**].

If $\mathbf{A} = \{[a_0, a_1, a_2, \ldots]; \ a_i = 1 \text{ or } 2\}$ is the set of numbers whose continued fraction representation contains only 1s and 2s then $\mathbf{A}$ is uncountable and so must contain transcendental numbers. However if $\alpha = [a_0, a_1, a_2, \ldots] \in \mathbf{A}$ then α has no close rational approximation. To see this let p_m/q_m be the mth convergent. For any simple continued fraction we have

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{1}{2}\left|\frac{p_{m+1}}{q_{m+1}} - \frac{p_m}{q_m}\right| = \frac{1}{2q_m q_{m+1}}.$$

But $q_{m+1} = q_m + q_{m-1}$ or $q_{m+1} = 2q_m + q_{m-1}$ since $a_m = 1$ or 2 so, for the numbers under consideration, we have $q_{m+1} \leqslant 3q_m$ and so

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{1}{6q_m^2}.$$

Moreover, this inequality holds for any rational number p/q, for if $|\alpha - p/q| \leqslant 1/6q^2$ then $|\alpha - p/q| < 1/2q^2$, and this implies that p/q is a convergent; see [**4**, p. 219].

REFERENCES

1. H. Davenport, The Higher Arithmetic, Dover, New York, 1983.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, London, 1960.
3. Hua Loo-Keng, Introduction to Number Theory, Springer, New York, 1982.
4. H. M. Stark, An Introduction to Number Theory, MIT Press, Cambridge, Massachusetts, 1981.

Visualizing the p-adic Integers

ALBERT A. CUOCO*
Mathematics Department, Woburn High School, Woburn, MA 01801

Algebra and analysis in the p-adic integers provide us with examples of mathematical phenomena that are hard to find in one package elsewhere. $\mathbb{Z}_p$ is a regular local ring; it is uncountable, and the non-archimedean valuation induced by its maximal ideal, $p\mathbb{Z}_p$, makes it into a compact, totally disconnected topological ring that contains the ordinary integers as a dense subring. How can one visualize such a creature?

Pictures that inspire images of p-adic objects are rare (see the frontispiece of [3] for an example). In this note, we show how to represent, for any prime p, the elements of $\mathbb{Z}_p$ as the points on a fractal in $\mathbb{R}^2$. The resulting correspondence captures many of the important algebraic and topological facets of $\mathbb{Z}_p$.

The pictures. Computer graphics enthusiasts have discovered a wonderful pastime. The idea is to plot an inductively defined (nondeterministic) sequence of points $\{Q_n\}$ in the plane. The sequence depends on an initial set of fixed vertices $\{A_0, A_1, \ldots, A_{k-1}\}$, the initial point Q_0 in the sequence and a real number e between 0 and 1. Suppose that Q_0 is one of the A_i and that Q_j has been defined. To find Q_{j+1}, pick one of the fixed vertices, say A_h, at random (that is, so that the probability of choosing any one vertex is $1/k$), and let δ be the distance between Q_j and A_h. Then Q_{j+1} is the point on the segment between Q_j and A_h at a distance of $e\delta$ from A_h. If $k = 3$, $e = .5$, and the A_i are vertices of an equilateral triangle, then the Q_n lie on a Sierpiński triangle (see [1] for the definition of Sierpiński triangles and Figure 1 for a picture of one).

We'll be concerned with the figures that occur when $k = p$ (a prime $\geqslant 3$), $e = 1/p$, and the A_i are vertices of a regular p-gon of side length 1. Call such a set of points S_p; a Macintosh computer running for a few minutes produced the approximations to S_3 in Figures 2a and 2b.

Figure 2a shows the process after 50 and 100 iterations; Figure 2b shows the results of 500 and 1000 iterations. Figure 2 shows the general shape of S_3; our definition of S_3 implies that it is the attractor for a system of three mappings on $\mathbb{R}^2$ (each mapping is a contraction with magnitude $1/3$ toward one of the A_i).[1]

A deterministic method to obtain S_p is to approximate it with sets $S_{p,n}$. Suppose, for example, that $p = 3$. Then $S_{3,0}$, $S_{3,1}$, $S_{3,2}$, $S_{3,3}$ and $S_{3,4}$ are depicted in Figure 3. Here, $S_{3,n}$ contains 3^n triangular regions. It is constructed from $S_{3,n-1}$ by replacing each triangular region T in $S_{3,n-1}$ with three small triangles, one from each corner of T, each with side-length $1/3$ that of T. S_3 is then the intersection of all the $S_{3,n}$. Similarly, S_p is the intersection of all the $S_{p,n}$. A few of the $S_{5,n}$ are illustrated in Figure 4.

Using this method for constructing S_p shows that S_p, as a subset of $\mathbb{R}^2$, is compact and totally disconnected.

*Work supported by NSF Grant MDR-895-4647

[1]To generate the images in Figure 2, the initial point Q_0 is chosen to be one of the vertices $\{A_0, A_1, A_2\}$; if Q_0 is not a vertex, the resulting sequence $\{Q_n\}$ will converge to S_3 in the sense that $\lim_{n\to\infty}(\inf_{P\in S_3} dist(P, Q_n)) = 0$.

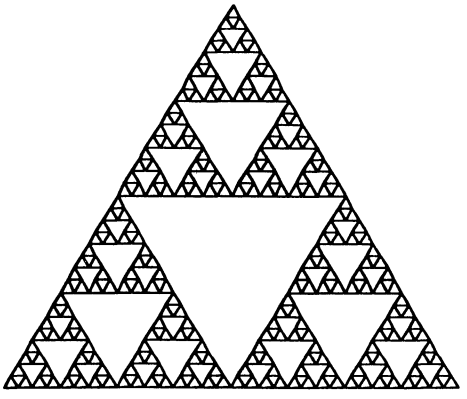


FIG. 1.

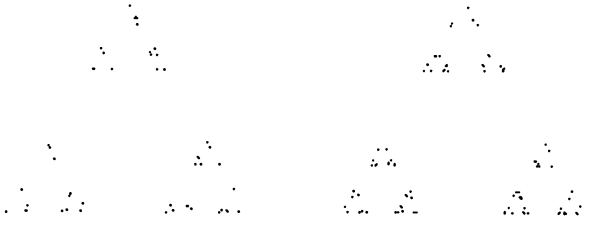


FIG. 2A.

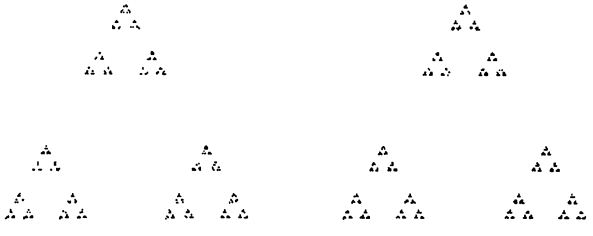


FIG. 2B.

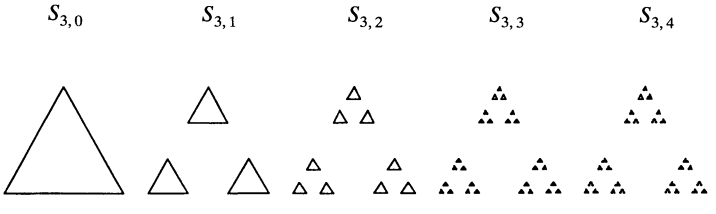


FIG. 3.

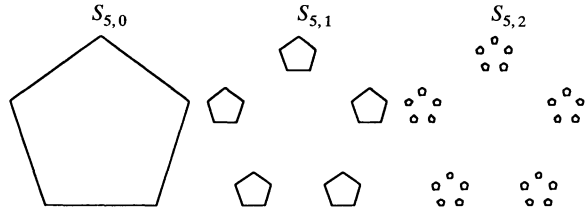


FIG. 4.

***p*-adic integers.** For precise definitions of $\mathbb{Z}_p$, see [3], [4] or [2]. We'll think of $\mathbb{Z}_p$ as the completion of $\mathbb{Z}$ with respect to the absolute value $|\ |_p$. Here,

$$|n|_p = p^{-\mathrm{ord}_p(n)}$$

where $\mathrm{ord}_p(n)$ is the highest exponent to which p divides n. The idea is that two integers are close if their difference is 0 modulo a high power of p. The completion contains the subring of $\mathbb{Q}$ known as the "p-integral" numbers (rational numbers whose denominators aren't divisible by p), as well as some irrational numbers.

If m is a positive integer, m has a finite base p expansion

$$m = m_0 + m_1 p + m_2 p^2 + m_3 p^3 + \cdots + m_r p^r,$$

where the m_i are integers between 0 and $p - 1$. This expansion (the *p-adic expansion* for m) will be denoted by its digits, and we'll write

$$m = m_0 m_1 m_2 m_3 \cdots m_r.$$

It's easy to see that $\mathrm{ord}_p(m)$ is the smallest integer k such that $m_k > 0$. It follows that two integers are close if their p-adic expansions agree for many places. In particular, the sequence

$$1,11,111,1111,11111,\ldots$$

is Cauchy, and its limit in $\mathbb{Z}_p$ can be calculated from the usual formula for the limit of a convergent geometric series

$$1111111\cdots = 1 + p + p^2 + p^3 + p^4 = \cdots = \frac{1}{1-p}.$$

(Notice that the common ratio in this series is p, and $|p|_p = 1/p$.) Since every p-adic integer is the limit of some Cauchy sequence of integers, and since the p-adic expansions for these integers agree for arbitrarily long initial strings, we can think of a p-adic integer as an infinite p-adic expansion, denoted by an infinite string of digits

$$a_0 a_1 a_2 a_3 a_4 \cdots = a_0 + a_1 p + a_2 p^2 + a_3 p^3 + a_4 p^4 + \cdots,$$

where each a_i is an integer between 0 and $p - 1$. As with all completions, $|\ |_p$ extends to a valuation on $\mathbb{Z}_p$, and its value can be calculated by the same formula that defines it on $\mathbb{Z}$ ($\mathrm{ord}_p(a)$ is the smallest integer k such that $a_k > 0$). So, just as in $\mathbb{Z}$, two p-adic integers are close if their representations agree for many digits (that is, if their difference is 0 modulo a high power of p). An element of $\mathbb{Z}_p$ is a non-negative integer if and only if its digits are eventually 0; an element of $\mathbb{Z}_p$ is in $\mathbb{Q}$ precisely when its digits eventually repeat.

The arithmetic of $\mathbb{Z}_p$ is especially simple. $\mathbb{Z}_p$ has unique factorization, and the only prime is p. Every p-adic integer a can be written uniquely as $p^e u$. Here, $e = \operatorname{ord}_p(a)$, and u is a unit (a p-adic integer whose first digit is not 0). $\mathbb{Z}_p/p\mathbb{Z}_p$ is isomorphic with $\mathbb{Z}/p\mathbb{Z}$. If U is the unit group in $\mathbb{Z}_p$ and U_1 is the subgroup of U consisting of units whose first digit is 1 (the *principal* units), then U/U_1 is isomorphic with $(\mathbb{Z}/p\mathbb{Z})^*$, the multiplicative group of non-zero elements in $\mathbb{Z}/p\mathbb{Z}$ (a cyclic group of order $p-1$).

Given a p-adic integer a, a fundamental system of neighborhoods for a is the sequence

$$\{a + p^n\mathbb{Z}_p : n = 0, 1, 2, 3, \ldots\}.$$

Indeed, given an integer n, $\mathbb{Z}_p$ splits up into p^n disjoint disks of diameter $1/p^n$, namely the cosets in $\mathbb{Z}_p/p^n\mathbb{Z}_p$. Two p-adic integers x and y are within $1/p^n$ of each other if and only if they belong to the same disk.

The metric d defined by $|\ |_p$ satisfies a stronger condition than the triangle inequality; if a, b and c are in $\mathbb{Z}_p$, then

$$d(a,b) \leqslant \max\{d(a,c), d(b,c)\}$$

(equality holds if $d(a,c)$ and $d(b,c)$ are unequal). This *non-archimedean* property of d implies that every triangle is isosceles and that every point interior to a circle is its center.

The correspondence. Naïvely, a p-adic integer is an object that can be approximated modulo p^n (that is, within a p-adic distance of $1/p^n$), for any n, by a length n p-adic expansion of an ordinary integer. And, a point on S_p is a point in $\mathbb{R}^2$ which, for any n, is in a connected component of $S_{p,n}$ (so that it can be approximated within a (Euclidean) distance of $1/p^n$ by a vertex on $S_{p,n}$). So, we can build a natural bijection between $\mathbb{Z}_p$ and S_p by setting up, for each n, a bijection between $\mathbb{Z}_p/p^n\mathbb{Z}_p$ and the vertices of $S_{p,n}$ that makes the following diagram commute

$$\begin{array}{ccc} \mathbb{Z}_p/p^{n+1}\mathbb{Z}_p & \longrightarrow & \mathbb{Z}_p/p^n\mathbb{Z}_p \\ \downarrow & & \downarrow \\ S_{p,n+1} & \longrightarrow & S_{p,n} \end{array}$$

Here, the top row is the usual homomorphism (it amounts to forgetting the last digit in the p-adic expansion), and the bottom row collapses each polygon of side length $1/p^{n+1}$ onto the unique vertex of the polygon of side length $1/p^n$ that it contains. Taking the inverse limit, we have a correspondence between $\mathbb{Z}_p$ and S_p.

Suppose, for the sake of example, that $p = 3$. Figure 5 shows how the correspondence develops.

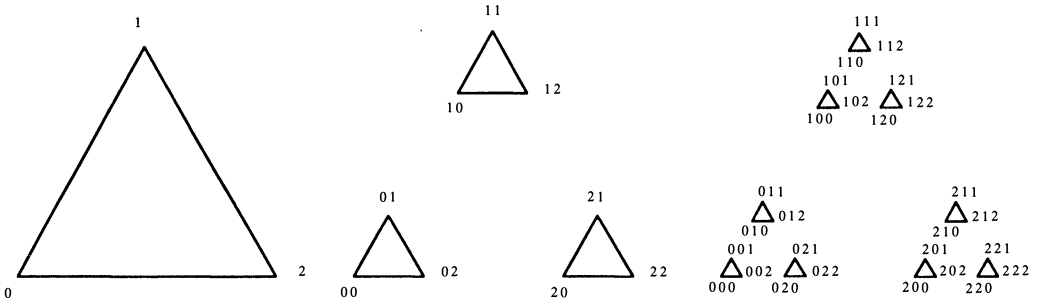


FIG. 5.

Figure 6 shows the skeleton of $S_{3,3}$, but the labeled points are the actual 3-adic integers (expressed as elements of $\mathbb{Q}$) that remain after passing to the limit.

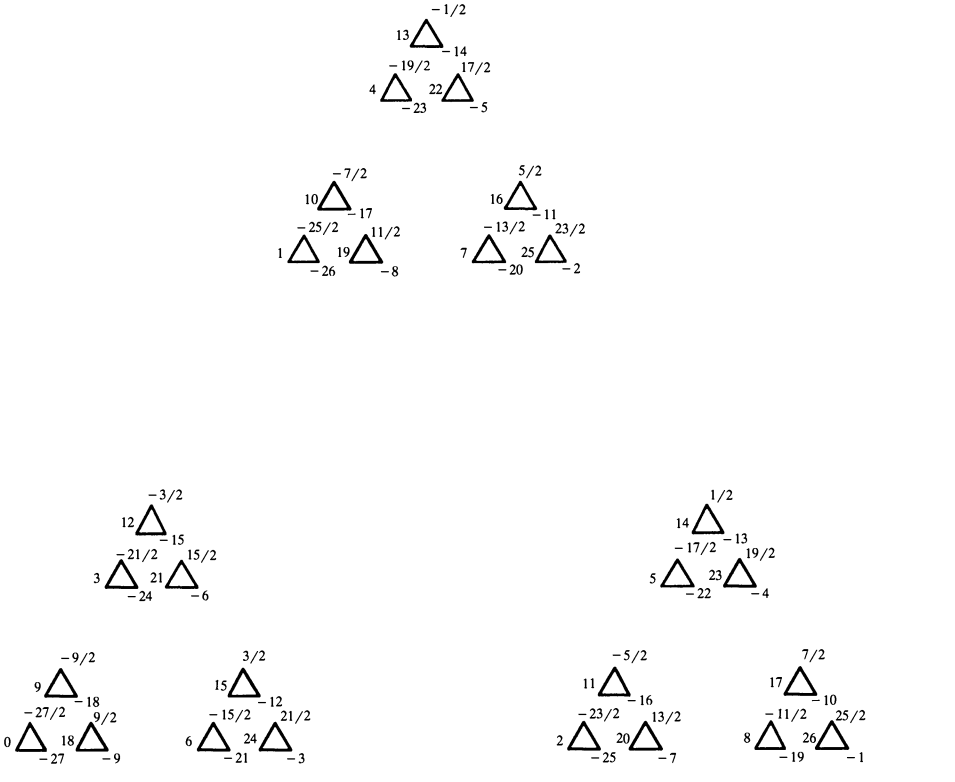


FIG. 6.

Similarly, Figure 7 shows some elements of $\mathbb{Z}_5$ as they appear on $S_{5,1}$.

Figure 8 shows how to chase down the (irrational) 5-adic integer 0313131111311113111113... in S_5 by tracking the approximation 0, 03, 031, and 0313.

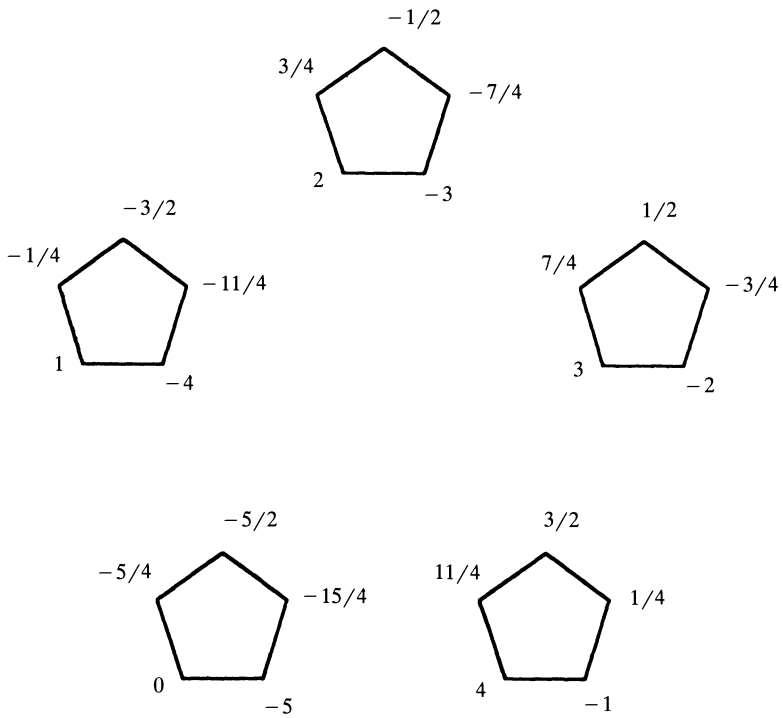


FIG. 7.

The most important aspect of our construction is: *Two p-adic integers are in a disk of diameter* $1/p^n$ *if and only if their images in* S_p *are in the same connected component of* $S_{p,n}$.

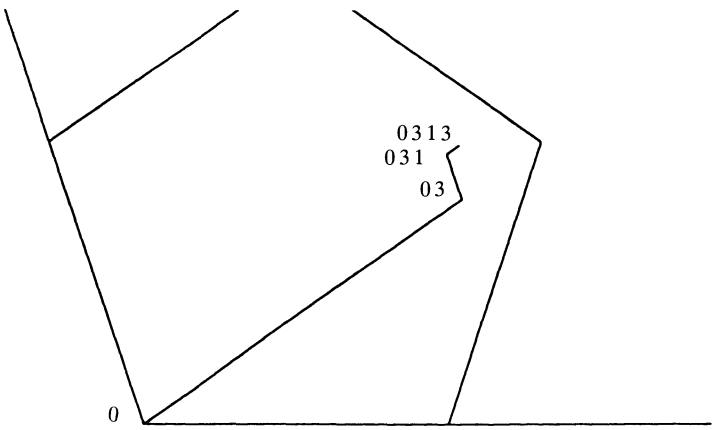


FIG. 8.

Using the correspondence. Looking at Figure 6, for example, the following facts about $\mathbb{Z}_3$ stand out:

1. $\mathbb{Z}_3$ is compact.

2. Every point in $\mathbb{Z}_3$ has a neighborhood that is both open and closed; $\mathbb{Z}_3$ is totally disconnected.

3. The image of the filtration $\{3^n\mathbb{Z}_3 : n = 0, 1, 2, 3, \dots\}$ in S_3 is contained in the set of triangles that contain 0 and have sides of length $(1/3^n)$ $(n = 0, 1, 2, \dots)$.

4. For any n, there are precisely 3^n disks of diameter $1/3^n$. These are the translates of $3^n\mathbb{Z}_3$ by $0, 1, 2, \dots, 3^{n-1}$.

Figure 9 suggests that $\mathbb{Z}_p/p\mathbb{Z}_p$ identifies with $\mathbb{Z}/p\mathbb{Z}$, and that U/U_1 identifies with $(\mathbb{Z}/p\mathbb{Z})^*$.

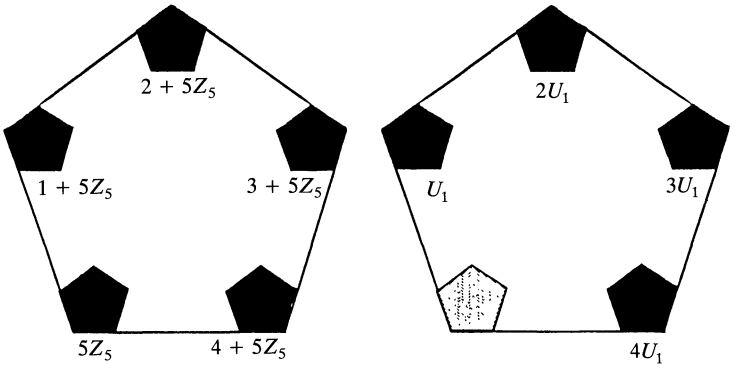


FIG. 9.

One can even use S_p to give examples of the non-archimedean geometry of $\mathbb{Z}_p$. For example, S_5 can be used to illustrate the isosceles triangle property in $\mathbb{Z}_5$. Suppose that we have three points A, B, and C in $\mathbb{Z}_5$ that are not the vertices of an equilateral triangle. Suppose further that

$$\max\{d(A,B), d(B,C), d(A,C)\} = \frac{1}{5^n}$$

and that this maximum distance is between A and C. Then one pair of points, say, A and B, are separated by a distance of at most $1/5^{n+1}$. Using our correspondence, the situation is as in Figure 10.

In Figure 10, we have a disk of diameter $1/5^n$ in $\mathbb{Z}_5$ that contains A, B, and C. This disk is the union of 5 disjoint disks, each of diameter $1/5^{n+1}$. Two elements of the large disk are in the same small disk if and only if the distance between them is at most $1/5^{n+1}$, while two 5-adic integers from the large disk are in *different* small disks if and only if the distance between them is exactly $1/5^n$. It follows that A and B are in the same small disk, and that C is in a different small disk. Therefore,

$$d(A,C) = d(B,C) = \frac{1}{5^n}$$

and our triangle is isosceles.

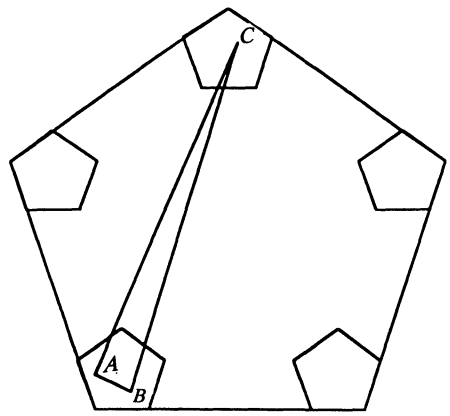


FIG. 10.

Conversely, the arithmetic of $\mathbb{Z}_p$ might be a useful tool in the study of the geometry of S_p. For example, suppose that c is a fixed integer between 1 and $p-1$. The fact that the neighborhoods of c are just the translates by c of the neighborhoods of 0 can be used to picture the effect of the mapping $x \mapsto c + x$ on S_p. Multiplication by c is a one-to-one mapping of $\mathbb{Z}_p$ onto itself, and there is an algorithm (much like the multiplication algorithm that is taught in elementary school) for obtaining the digits of ca from those of a. Given a p-adic integer a and its location in S_p, can this algorithm be transported to a simple geometric construction that allows one to locate ca in S_p?

Remark. In our definition of the p-adic absolute value

$$|a|_p = \left(\frac{1}{p}\right)^{\operatorname{ord}_p(a)},$$

our choice of the "uniformizing parameter" $1/p$ is somewhat arbitrary. Any real number less that 1 will give an equivalent metric, although our correspondence requires a parameter less than 0.5. In the case $p = 2$, it is sometimes convenient to take $1/4$, so that

$$|a|_2 = \left(\frac{1}{4}\right)^{\operatorname{ord}_2(a)}.$$

Then it is possible to set up our correspondence between $\mathbb{Z}_2$ and the Cantor-like subset of the unit interval obtained by repeatedly removing the middle half.

***p*-adic numbers.** Because $\mathbb{Z}_p$ is an integral domain, it has a quotient field $\mathbb{Q}_p$, the field of p-adic numbers. $\mathbb{Q}_p$ is locally compact, and it contains $\mathbb{Z}_p$ as an open subring. If $|\ |_p$ is viewed as a function on $\mathbb{Q}$ (by defining the absolute value of a quotient as the quotient of the absolute values), $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to $|\ |_p$. We end by sketching a method for extending our embedding of $\mathbb{Z}_p$ into $\mathbb{R}^2$ to an embedding of $\mathbb{Q}_p$ into $\mathbb{R}^2$.

One way to think about a p-adic number is as a p-adic expansion that contains a finite number of terms with negative exponents. If a is the p-adic number

$$a = \sum_{k=-N}^{\infty} a_k p^k,$$

we'll denote a by its digits

$$a = a_{-N}a_{-N+1} \cdots a_{-1}.a_0a_1a_2 \ldots .$$

We can write $\mathbb{Q}_p$ as the increasing union of the sets $\mathbb{Q}_{p,n}$ where

$$\mathbb{Q}_{p,n} = \{\alpha \in \mathbb{Q}_p : |\alpha|_p \leqslant p^n\}.$$

Then

$$\mathbb{Q}_{p,0} = \mathbb{Z}_p = \{0.a_0a_1a_2 \cdots : 0 \leqslant a_i \leqslant p-1\},$$

$$\mathbb{Q}_{p,1} = \{a_{-1}.a_0a_1a_2 \cdots : 0 \leqslant a_i \leqslant p-1\},$$

and, more generally,

$$\mathbb{Q}_{p,n} = \{a_{-n} \cdots a_{-1}.a_0a_1a_2 \cdots : 0 \leqslant a_i \leqslant p-1\}.$$

Geometrically, we can replicate S_p as in Figure 11 (for $p = 3$) to obtain a fractal $S_{p,\infty}$.

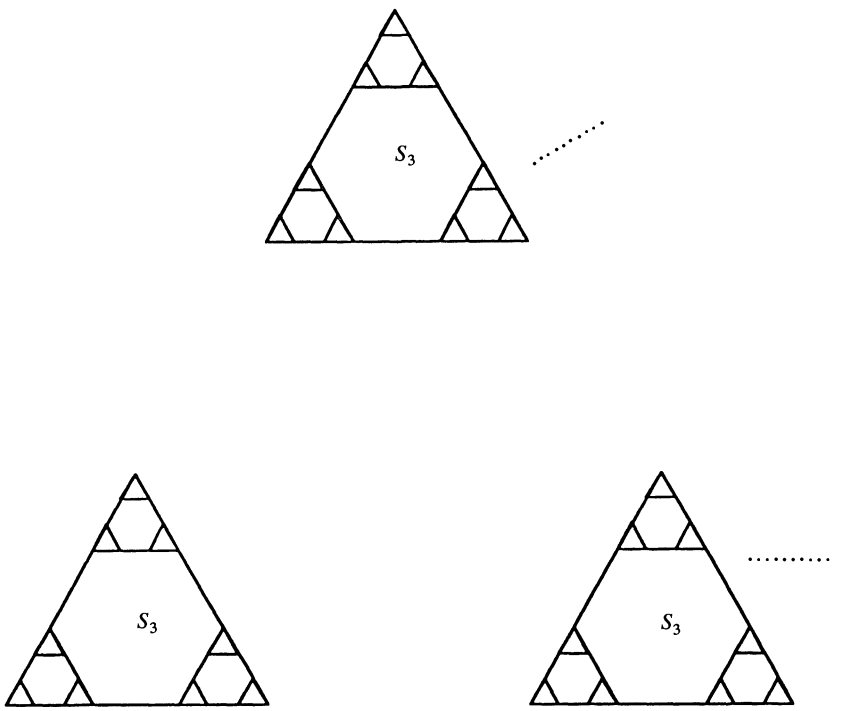


FIG. 11.

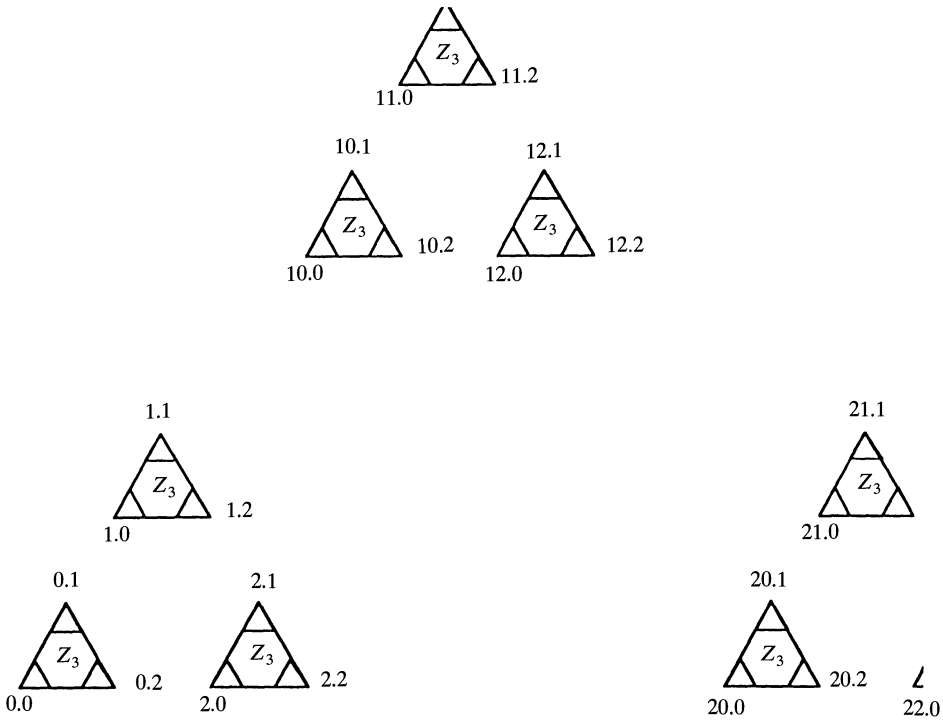


FIG. 12.

The bijection between $\mathbb{Q}_p$ and $S_{p,\infty}$ is shown in Figure 12 for the case $p = 3$.

Special thanks to E. Paul Goldenberg.

REFERENCES

1. Michael Barnsley, Fractals Everywhere, Academic Press, San Diego, 1988.
2. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.
3. Neal Koblitz, p-adic Numbers, p-adic Analysis, and Zeta-Functions, Springer Verlag, New York, 1977.
4. Jean-Pierre Serre, A Course in Arithmetic, Springer Verlag, New York, 1973.

Editor's note: Several readers have pointed out that the argument presented in Deng Bo's article "The derivation of the Maclaurin series for sine and cosine" in November 1990 issue of the MONTHLY is well known. It appears in some calculus books, including those by Ivan Niven and George Simmons. Niven points out that the argument appears in the 1941 book *What is Mathematics?*, by R. Courant and H. Robbins, but was probably well known several decades earlier. Finally, Niven's book (*Calculus: An Introductory Approach*, 2d edition, Von Nostrand, New York, 1966, pp. 140, 160, 167, 171) shows how the logarithmic, exponential, and arctangent series can all be given similar derivations.

us call an integer *good* if its odd part is less than or equal to n. There are $\lfloor (n+1)/2 \rfloor$ odd integers in $[1, n]$ and, if k is any one of them, then it is easily seen that $n < 2^r k \leqslant 2^m n$ for exactly m integers r. Therefore the number of good integers in the interval $(n, 2^m n]$ is $m\lfloor (n+1)/2 \rfloor$. It follows that the set S contains at least

$$(2^m - 1)n + 1 - \left\{ 2^m n - n - m\left\lfloor \frac{n+1}{2} \right\rfloor \right\} = 1 + m\left\lfloor \frac{n+1}{2} \right\rfloor$$

good integers. Now, by the pigeonhole principle there must exist an odd k $(1 \leqslant k \leqslant n)$ such that S contains at least $m + 1$ integers with odd part k. If these integers are arranged in increasing order it is obvious that each of them divides the next one. (Moreover, all the quotients are powers of 2.)

The example $\{n+1, n+2, \ldots, 2^m n\}$ shows that the assertion of the problem need not hold for sets of cardinality $(2^m - 1)n$; for if $a_0, a_1, \ldots, a_m$ are distinct positive integers with $a_{i-1} | a_i$ for $i = 1, 2, \ldots, m$, then clearly $a_m \geqslant 2^m a_0$.

Editorial comment. Most other solvers also used the pigeonhole principle. The proposer proved the result by induction on n, P. G. Walsh and D. Vuignier proved the result by induction on m, while John Steinig gave two inductive proofs, one by induction on m and one by induction on n.

Several solvers mentioned the apparently more general result: If $N \geqslant 2^m n$ and if T is a subset of $\{1, 2, \ldots, N\}$ of cardinality $N - n + 1$, then T contains $m + 1$ distinct integers $a_0, a_1, \ldots, a_m$ such that $a_{i-1} | a_i$ for $i = 1, 2, \ldots, m$. This follows by applying the result of the problem to

$$S = T \cap \{1, 2, 3, \ldots, 2^m n\}.$$

Solved also by J. C. Binz (Switzerland), D. Callan, R. J. Chapman (England), L. E. Clarke (England), M. E. Kuczma (Poland), P. S. Livingston, O. P. Lossers (The Netherlands), D. E. Manes, A. Nijenhuis, E. R. Scheinerman, A. J. Schwenk, J. H. Steelman, J. Steinig (Switzerland), P. G. Walsh & D. Vuignier (Canada), A. Zulauf (New Zealand), and the proposer.

Sparse Subsets of $\mathbb{Z}$ mod $4n$

E 3328 [1989, 446]. *Proposed by Bruce K. Smith, San Rafael, CA and James G. Propp, Berkeley, CA.*

Suppose n is a positive integer greater than one. Call a subset S of $\mathbb{Z} \bmod 4n$ *sparse* if for every k in $\mathbb{Z} \bmod n$ the following two conditions are satisfied:

(a) $S \cap \{4k - 2, 4k - 1, 4k, 4k + 1, 4k + 2\}$ has at most two elements,

(b) $S \cap \{4k + 1, 4k + 2, 4k + 3\}$ has at most one element.

Prove that $\mathbb{Z} \bmod 4n$ has 7^n sparse subsets.

Solution by Raphael M. Robinson, University of California, Berkeley. Call a subset S' of $\mathbb{Z} \bmod 4n$ *spare* if for every k in $\mathbb{Z} \bmod n$ the following two conditions are satisfied:

(a′) $S' \cap \{4k - 1, 4k, 4k + 1\}$ has at most two elements,

(b′) $S' \cap \{4k + 2\}$ is empty.

Since for each k in $\mathbb{Z} \bmod n$ there are precisely 7 choices for $S' \cap \{4k - 1, 4k, 4k + 1\}$, and the choices are independent for different values of k, there are precisely 7^n spare subsets of $\mathbb{Z} \bmod 4n$.

We establish a one-to-one correspondence between sparse subsets S and spare subsets S' as follows. Given a sparse subset S, we construct a spare subset S' by replacing each element in S of the form $4k + 2$ by both $4k + 1$ and $4k + 3$, which were absent in S. Then the resulting S' will be spare, since the count is the same in (a') as in (a). (We need $n > 1$ to make $4k - 2$ and $4k + 2$ distinct.) Given a spare subset S', we construct a sparse subset S by replacing each occurrence of both $4k + 1$ and $4k + 3$ in S' by $4k + 2$. The resulting set will be sparse, since the count in (a) is the same as the count in (a'), and the count in (b) is reduced to one. If at most one of the numbers $4k + 1$ and $4k + 3$ is in S', then (b) is satisfied without any change being made. It is easy to check that for $n > 1$ these two constructions give the desired one-to-one correspondence.

Solved also by J. C. Binz (Switzerland), R. J. Hendel, A. A. Jagers (The Netherlands), I. Kozma (Israel), M. E. Kuczma (Poland), O. P. Lossers (The Netherlands), L. F. Martins (student), A. Nijenhuis, J. H. Steelman, and the proposers.

Schröder Numbers Divisible by 3

E 3343 [1989, 734]. *Proposed by Lou Shapiro, Howard University, Washington, DC and Douglas Rogers, University of East Anglia, U.K.*

Let S_n be the number of polygonal paths in the plane that start at $(0, 0)$, end at (n, n), contain no points above the line $y = x$, and are composed of steps taken from the set $\{(0, 1), (1, 0), (1, 1)\}$. For example, $S_0 = 1, S_1 = 2, S_2 = 6, S_3 = 22, S_4 = 90$. (The numbers in the sequence $\{S_n\}$ are often called the *Schröder numbers*.)

Prove that S_n is divisible by 3 for every positive even integer n.

Solution by Carl Schoen and Paolo Ranaldi, University of Wisconsin, Eau Claire. Let A_k be the number of such paths whose penultimate vertex on the diagonal $y = x$ is (k, k). Then $S_n = \sum_{k=0}^{n-1} A_k$. Then A_0 is the number of paths from $(1, 0)$ to $(n, n - 1)$ containing no points above the diagonal connecting $(1, 0)$ to $(n, n - 1)$. Hence $A_0 = S_{n-1}$. For $k = 1, 2, \ldots, n - 2$, A_k is the number of paths from $(0, 0)$ to (k, k) multiplied by the number of paths from $(k + 1, k)$ to $(n, n - 1)$ containing no points above the diagonal connecting $(k + 1, k)$ to $(n, n - 1)$. Hence, $A_k = S_k S_{n-1-k}$. Finally, A_{n-1} is the number of paths from $(0, 0)$ to $(n - 1, n - 1)$ multiplied by the number of paths from $(n - 1, n - 1)$ to (n, n), or $A_{n-1} = 2S_{n-1}$. It follows that $S_n = 3S_{n-1} + \sum_{k=1}^{n-2} S_k S_{n-1-k}$.

Assume S_k is divisible by 3 for all even $k < n$. If n is even, then each term in the above sum is divisible by 3, since $3|S_k$ when k is even and $3|S_{n-1-k}$ when k is odd. It follows by induction that $3|S_n$ for all even n.

Solved also by the proposers and 37 other readers.

Sum-free Sets Modulo *n*

E 3346 [1989, 735]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

Call a subset T of $\mathbb{Z} \bmod n$ *sum-free* if the sum of two distinct elements of T is not in T. Let $s(n)$ be the maximum cardinality of a sum-free subset of $\mathbb{Z} \bmod n$.

(a) Prove that $s(n) = n/2$ if n is even.
(b) If n is odd, prove that $s(n) \geqslant \lfloor n/3 \rfloor + 1$.
(c)* If n is odd, is $s(n) = \lfloor n/3 \rfloor + 1$?

Solution by Torleiv Kløve, Bergen, Norway. (a) If n is even, the set of odd integers is a sum-free set of size $n/2$. On the other hand, if T is sum-free and $a \in T$, let $S = \{a + b: b \in T, b \neq a\}$. Then S and T are disjoint, which implies $n \geqslant 2|T| - 1$, or $|T| \leqslant \lfloor (n + 1)/2 \rfloor = n/2$.

(b) The set $\{a: \lfloor n/3 \rfloor \leqslant a \leqslant 2\lfloor n/3 \rfloor\}$ is sum-free and has $\lfloor n/3 \rfloor + 1$ elements.

(c) The answer is "sometimes, but not always." Let p be a prime such that $p \equiv -1 (\mathrm{mod}\, 6)$, and let n be a multiple of p. Let T be the subset of $\mathbb{Z} \bmod n$ consisting of those residue classes $a \bmod n$ such that a is congruent modulo p to one of the $(p + 1)/3$ integers in the interval $[(p + 1)/3, (2p - 1)/3]$. If $c \equiv (a + b) (\mathrm{mod}\, n)$ for $a, b \in T$, then $c \equiv (a + b) (\mathrm{mod}\, p)$, which implies $c \notin T$; thus T is sum-free. Further $|T| = n(p + 1)/(3p)$. If $n = p$ or $n = 3p$, then $n(p + 1)/(3p) = \lfloor n/3 \rfloor + 1$; in these two cases presumably no larger sum-free set exists. However, if $n \geqslant 5p$, then $n(p + 1)/(3p) > \lfloor n/3 \rfloor + 1$.

Editorial comment. Concerning the correct answer for odd n, five solvers (including Kløve), conjectured that $s(n) = \lfloor n/3 \rfloor + 1$ if n has no prime divisors congruent to 5 mod 6, but that $s(n) = n(p + 1)/(3p)$ if p is the smallest prime divisor of n congruent to 5 mod 6. Kløve commented that exhaustive search verifies this for $n \leqslant 55$. Thus $n = 25$ is the smallest value of n for which $s(n) > \lfloor n/3 \rfloor + 1$.

This problem requires that the summands be distinct. The variation arising by dropping that requirement is considered in H. L. Abbott and E. T. H. Wang, "Sum-free sets of integers," *Proc. Am. Math. Soc.* 67(1977), 11–16.

Added in proof. In a manuscript received by the editors after the above material was sent to the printer, Umberto Zannier (Venice, Italy) proved the conjecture enunciated in the first paragraph of the Editorial comment. He used a theorem of Martin Kneser given as Theorem 16 of chapter 1 of H. Halberstam and K. F. Roth, *Sequences*, second edition, Springer-Verlag, 1983. Professor Zannier will publish his proof elsewhere.

Solved also by J. Duemmel, R. B. Eggleton (Brunei), F. Galvin, D. R. Guichard, O. P. Lossers (Netherlands), J. H. Nieto (Venezuela), R. B. Richter (Canada), J. H. Steelman, R. Stong, G. Torchinelli, B. M. de Weger (Netherlands), J. Zhang (China). Solutions to (a) and (b) only were submitted by F. Balogh (Hungary), G. E. Bilodeau, R. J. Hendel, W. Janous (Austria), R. Martin (student), and the proposer.

A Series for e

E 3352 [1989, 838]. *Proposed by G. I. Senum, Brookhaven National Laboratory, Upton, Long Island, NY.*

Show that

$$\sum_{n=0}^{\infty} \frac{1}{n!(n^4 + n^2 + 1)} = \frac{e}{2}.$$

(a) Prove that $s(n) = n/2$ if n is even.
(b) If n is odd, prove that $s(n) \geqslant \lfloor n/3 \rfloor + 1$.
(c)* If n is odd, is $s(n) = \lfloor n/3 \rfloor + 1$?

Solution by Torleiv Kløve, Bergen, Norway. (a) If n is even, the set of odd integers is a sum-free set of size $n/2$. On the other hand, if T is sum-free and $a \in T$, let $S = \{a + b: b \in T, b \neq a\}$. Then S and T are disjoint, which implies $n \geqslant 2|T| - 1$, or $|T| \leqslant \lfloor (n + 1)/2 \rfloor = n/2$.

(b) The set $\{a: \lfloor n/3 \rfloor \leqslant a \leqslant 2\lfloor n/3 \rfloor\}$ is sum-free and has $\lfloor n/3 \rfloor + 1$ elements.

(c) The answer is "sometimes, but not always." Let p be a prime such that $p \equiv -1 (\mathrm{mod}\, 6)$, and let n be a multiple of p. Let T be the subset of $\mathbb{Z} \bmod n$ consisting of those residue classes $a \bmod n$ such that a is congruent modulo p to one of the $(p + 1)/3$ integers in the interval $[(p + 1)/3, (2p - 1)/3]$. If $c \equiv (a + b) (\mathrm{mod}\, n)$ for $a, b \in T$, then $c \equiv (a + b) (\mathrm{mod}\, p)$, which implies $c \notin T$; thus T is sum-free. Further $|T| = n(p + 1)/(3p)$. If $n = p$ or $n = 3p$, then $n(p + 1)/(3p) = \lfloor n/3 \rfloor + 1$; in these two cases presumably no larger sum-free set exists. However, if $n \geqslant 5p$, then $n(p + 1)/(3p) > \lfloor n/3 \rfloor + 1$.

Editorial comment. Concerning the correct answer for odd n, five solvers (including Kløve), conjectured that $s(n) = \lfloor n/3 \rfloor + 1$ if n has no prime divisors congruent to 5 mod 6, but that $s(n) = n(p + 1)/(3p)$ if p is the smallest prime divisor of n congruent to 5 mod 6. Kløve commented that exhaustive search verifies this for $n \leqslant 55$. Thus $n = 25$ is the smallest value of n for which $s(n) > \lfloor n/3 \rfloor + 1$.

This problem requires that the summands be distinct. The variation arising by dropping that requirement is considered in H. L. Abbott and E. T. H. Wang, "Sum-free sets of integers," *Proc. Am. Math. Soc.* 67(1977), 11–16.

Added in proof. In a manuscript received by the editors after the above material was sent to the printer, Umberto Zannier (Venice, Italy) proved the conjecture enunciated in the first paragraph of the Editorial comment. He used a theorem of Martin Kneser given as Theorem 16 of chapter 1 of H. Halberstam and K. F. Roth, *Sequences*, second edition, Springer-Verlag, 1983. Professor Zannier will publish his proof elsewhere.

Solved also by J. Duemmel, R. B. Eggleton (Brunei), F. Galvin, D. R. Guichard, O. P. Lossers (Netherlands), J. H. Nieto (Venezuela), R. B. Richter (Canada), J. H. Steelman, R. Stong, G. Torchinelli, B. M. de Weger (Netherlands), J. Zhang (China). Solutions to (a) and (b) only were submitted by F. Balogh (Hungary), G. E. Bilodeau, R. J. Hendel, W. Janous (Austria), R. Martin (student), and the proposer.

A Series for e

E 3352 [1989, 838]. *Proposed by G. I. Senum, Brookhaven National Laboratory, Upton, Long Island, NY.*

Show that

$$\sum_{n=0}^{\infty} \frac{1}{n!(n^4 + n^2 + 1)} = \frac{e}{2}.$$

Prime Factors of $1 + 9^a$

E 3355 [1989, 927]. *Proposed by Leo J. Alex, State University of New York, Oneonta.*

Characterize those positive integers a such that the prime factors of $1 + 9^a$ are all less than 40. Some useful factorizations:

$$1 + 9^4 = 2 \cdot 17 \cdot 193, \qquad 1 + 9^5 = 2 \cdot 5^2 \cdot 1181,$$
$$1 + 9^6 = 2 \cdot 41 \cdot 6481, \qquad 1 + 9^7 = 2 \cdot 5 \cdot 29 \cdot 16493.$$

Solution (independently) by Hugh Noland, Chico, CA, by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA, and by Richard Stong, UCLA. The only such positive integer is 1. If p is an odd prime dividing $1 + 9^a$, then -1 is a quadratic residue of p (since 9 is a square), and hence $p \equiv 1 \pmod 4$. The only such primes less than 40 are $\{5, 13, 17, 29, 37\}$. Modulo 13, the powers of 9 are $\{1, 3, 9\}$, and modulo 37 they are $\{1, 7, 9, 10, 12, 16, 26, 33, 34\}$ so neither of these primes ever divides $1 + 9^a$. Hence the only possible prime factors less than 40 are $\{2, 5, 17, 29\}$.

Consider the powers of 9 modulo $17, 25, 29$. If 17 divides $1 + 9^a$, then $a \equiv 4 \pmod 8$ and hence 9^a is an odd power of a number 9^4 that is congruent to $-1 \pmod{193}$, as given above, so 193 divides $1 + 9^a$. Similarly, if 25 divides $1 + 9^a$, then $a \equiv 5 \pmod{10}$ and 1181 also divides $1 + 9^a$, and if 29 divides $1 + 9^a$, then $a \equiv 7 \pmod{14}$ and 16493 also divides $1 + 9^a$. Finally, since $9 \equiv 1 \pmod 4$, we have $1 + 9^a \equiv 2 \pmod 4$, so 4 never divides $1 + 9^a$. Hence if all prime factors of $1 + 9^a$ are less than 40, then we have at most one power of 2, at most one power of 5, and no powers of 17 or 29, so $1 + 9^a = 10$.

Editorial comment. With a little extra work, B. M. M. de Weger showed that only for $a = 1, 2, 3, 4$ are all the prime divisors of $1 + 9^a$ less than 200.

Twenty-five correct solutions and one incorrect solution were received.

Limit of a Reciprocal Sequence

E 3356 [1989, 927]. *Proposed by Wu Wei Chao, Jinhua City, Zhejiang, China.*

Suppose $a_1 = 1$ and $a_{n+1} = n/a_n$ for $n = 1, 2, \ldots$. Evaluate

$$\lim_{n\to\infty} n^{-1/2}\left\{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right\}.$$

Solution dedicated to the late Michael J. Dixon by his colleagues in the Chico Problem Group, California State University, Chico. We show that the limit is $\sqrt{2/\pi} + \sqrt{\pi/2}$. Using the recurrence, we have $a_{n+1} \cdot a_n = n$ and $a_n \cdot a_{n-1} = n - 1$. Subtracting these yields $a_{n+1} - a_{n-1} = a_n^{-1}$, so the sum of reciprocals telescopes to yield $a_2^{-1} + \cdots + a_n^{-1} = a_{n+1} + a_n - a_1 - a_2$. Hence

$$\lim_{n\to\infty} \frac{a_1^{-1} + \cdots + a_n^{-1}}{\sqrt{n}} = \lim_{n\to\infty}\left(\frac{a_{n+1}}{\sqrt{n}} + \frac{a_n}{\sqrt{n}}\right). \tag{1}$$

Since $a_{n+1} = n/a_n = a_{n-1}n/(n-1)$, we can iterate to obtain

$$a_{2n+2} = \frac{(2n+1)!}{2^{2n}(n!)^2}, \qquad a_{2n+1} = \frac{2^{2n}(n!)^2}{(2n)!}.$$

By Stirling's formula, we have $a_{2n+2} \sim \sqrt{2/\pi} \cdot \sqrt{2n}$ and $a_{2n+1} \sim \sqrt{\pi/2} \cdot \sqrt{2n}$. Hence the two contributions to the right-hand side of (1) approach $\sqrt{2/\pi}$ and $\sqrt{\pi/2}$.

Fifty-seven correct solutions and one incorrect solution were received.

ADVANCED PROBLEMS

6655. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

Given a positive integer n, we wish to find distinct integers greater than n such that their product with n is a square and we wish to do this in such a way that the largest number used is as small as possible. Let $f(n)$ be this minimal value of the largest number used. For example $f(10) = 18$, since $10 \cdot 12 \cdot 15 \cdot 18 = (180)^2$ and the product of 10 with any subset of $\{11, 12, 13, 14, 15, 16, 17\}$ is not a square.

Show that if $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ is any sequence of positive numbers tending to zero, then the set

$$\{n : f(n) - n \geqslant n^{1-\varepsilon_n}\}$$

has density zero.

6656. *Proposed by Dragomir Ž. Đoković, University of Waterloo, Ontario.*

(i) If $P(z)$ is a non-zero polynomial over $\mathbb{C}$ with at least two distinct zeros, prove that $1/P(z)$ has at least one non-zero residue (or equivalently $1/P(z)$ is not the derivative of a rational function).

(ii) Prove or disprove the following more general assertion: If f and g are rational functions over $\mathbb{C}$ such that $f'(z)g'(z) = 1$, then $f'(z) = c(z-a)^k$ for some complex constants c, a and some integer $k \neq \pm 1$.

6657. *Proposed by J. J. Rotman and P. M. Weichsel, University of Illinois at Urbana-Champaign.*

Suppose we make a finite group G into a graph Γ by defining two elements a, b of G to be adjacent if $(ab^{-1})^2 \neq e$. If the elements of G are $v_1, v_2, \ldots, v_n$, let A be the n by n matrix in which the element in the ith row and jth column is 1 if v_i and v_j are adjacent and is 0 if v_i and v_j are not adjacent. (A is an adjacency matrix for Γ.)

Show that $\det A$ is an even integer.

Since $a_{n+1} = n/a_n = a_{n-1}n/(n-1)$, we can iterate to obtain

$$a_{2n+2} = \frac{(2n+1)!}{2^{2n}(n!)^2}, \qquad a_{2n+1} = \frac{2^{2n}(n!)^2}{(2n)!}.$$

By Stirling's formula, we have $a_{2n+2} \sim \sqrt{2/\pi} \cdot \sqrt{2n}$ and $a_{2n+1} \sim \sqrt{\pi/2} \cdot \sqrt{2n}$. Hence the two contributions to the right-hand side of (1) approach $\sqrt{2/\pi}$ and $\sqrt{\pi/2}$.

Fifty-seven correct solutions and one incorrect solution were received.

ADVANCED PROBLEMS

6655. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

Given a positive integer n, we wish to find distinct integers greater than n such that their product with n is a square and we wish to do this in such a way that the largest number used is as small as possible. Let $f(n)$ be this minimal value of the largest number used. For example $f(10) = 18$, since $10 \cdot 12 \cdot 15 \cdot 18 = (180)^2$ and the product of 10 with any subset of $\{11, 12, 13, 14, 15, 16, 17\}$ is not a square.

Show that if $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ is any sequence of positive numbers tending to zero, then the set

$$\{n : f(n) - n \geqslant n^{1-\varepsilon_n}\}$$

has density zero.

6656. *Proposed by Dragomir Ž. Đoković, University of Waterloo, Ontario.*

(i) If $P(z)$ is a non-zero polynomial over $\mathbb{C}$ with at least two distinct zeros, prove that $1/P(z)$ has at least one non-zero residue (or equivalently $1/P(z)$ is not the derivative of a rational function).

(ii) Prove or disprove the following more general assertion: If f and g are rational functions over $\mathbb{C}$ such that $f'(z)g'(z) = 1$, then $f'(z) = c(z-a)^k$ for some complex constants c, a and some integer $k \neq \pm 1$.

6657. *Proposed by J. J. Rotman and P. M. Weichsel, University of Illinois at Urbana-Champaign.*

Suppose we make a finite group G into a graph Γ by defining two elements a, b of G to be adjacent if $(ab^{-1})^2 \neq e$. If the elements of G are $v_1, v_2, \ldots, v_n$, let A be the n by n matrix in which the element in the ith row and jth column is 1 if v_i and v_j are adjacent and is 0 if v_i and v_j are not adjacent. (A is an adjacency matrix for Γ.)

Show that $\det A$ is an even integer.

SOLUTIONS OF ADVANCED PROBLEMS

Poles of Green's Function

6513 [1986, 218]. *Proposed by Maurice Machover, St. John's University, Jamaica, NY.*

(a) Show that Green's function for the operator $Lu = u''$, with any boundary conditions, has no pole of order higher than three.

(b) Devise boundary conditions (the simpler the better) for which there is a third order pole.

Solution by the proposer. We consider the Green's function for $u'' + \lambda^2 u$ on $[0, \pi]$ with fixed boundary conditions, and examine its properties as a meromorphic function of λ^2 at $\lambda^2 = 0$. That is, for $t \in [0, \pi]$ and fixed $\lambda \in \mathbb{C}$ we consider the function $G(x, t, \lambda)$, if indeed it exists, with the following properties: It satisfies the differential equation

$$G_{xx}(x, t, \lambda) + \lambda^2 G(x, t, \lambda) = \delta(x - t),$$

and the linear boundary conditions

$$B_1(G) := a_{11}G(0, t, \lambda) + a_{12}G_x(0, t, \lambda) + a_{13}G(\pi, t, \lambda) + a_{14}G_x(\pi, t, \lambda) = 0,$$
$$B_2(G) := a_{21}G(0, t, \lambda) + a_{22}G_x(0, t, \lambda) + a_{23}G(\pi, t, \lambda) + a_{24}G_x(\pi, t, \lambda) = 0,$$

where $a_{11}, \ldots, a_{24}$ are constants and B_1, B_2 the corresponding operators. We make the standard assumption that the matrix (a_{ij}) has rank two. Here $\delta(x - t)$ denotes the usual delta function.

Now $u_p = u_p(x, t, \lambda) = H(x - t)\lambda^{-1} \sin \lambda(x - t)$, where $H(x)$ is the usual Heaviside function, is a particular solution of the differential equation. Hence Green's function has the form

$$G(x, t, \lambda) = a(t, \lambda)\cos \lambda x + b(t, \lambda)\sin \lambda x + u_p(x, t, \lambda).$$

To determine $a = a(t, \lambda)$ and $b = b(t, \lambda)$ note that

$$aB_1(\cos \lambda x) + bB_1(\sin \lambda x) + B_1 u_p = 0$$
$$aB_2(\cos \lambda x) + bB_2(\sin \lambda x) + B_2 u_p = 0,$$

so by Cramer's rule $a = \Delta_1(t, \lambda)/\Delta(\lambda)$ and $b = \Delta_2(t, \lambda)/\Delta(\lambda)$, where

$$\Delta(\lambda) = \begin{vmatrix} B_1(\cos \lambda x) & B_1(\sin \lambda x) \\ B_2(\cos \lambda x) & B_2(\sin \lambda x) \end{vmatrix},$$

and similar formulas give Δ_1 and Δ_2. Note that u_p is analytic in λ at $\lambda = 0$, so any pole at $\lambda = 0$ must arise from a zero of $\Delta(\lambda)$ at $\lambda = 0$.

For part (*a*) observe that G cannot exist if $\Delta(\lambda)$ vanishes identically. Write

$$\Delta(\lambda) = \sum_{i=0}^{\infty} c_i \lambda^i.$$

The existence of a pole of order at least four in λ^2 implies that $c_i = 0$ for $i \leqslant 7$.

and

$$\Delta_2(t,\lambda) = \left[1 + \cos\lambda\pi - \frac{\lambda\pi}{15}\sin\lambda\pi\right]\left[-23\lambda^{-1}\sin\lambda(\pi - t)\right] + \left[\lambda^{-1}\sin\lambda(\pi - t) + \frac{\pi}{15}\cos\lambda(\pi - t)\right][22 + 23\cos\lambda\pi].$$

Since $\Delta_1(0) = 0$ we have $N(0) = 0$. Furthermore, as $\lambda \to 0$ the first term of $N(\lambda)$ is

$$K_1\lambda + O(\lambda^2)$$

where $K_1 \neq 0$ is *independent* of x, while the second term of $N(\lambda)$ is

$$K_2\lambda + O(\lambda^2),$$

where

$$K_2 = 2x[23(t - \pi)] + 45x\left[\pi - t + \frac{\pi}{15}\right] = (t + 2\pi)x.$$

Hence the zero here is of first order, and the result follows.

Editorial comment. The proposer evidently intended the underlying space to be an interval and the boundary conditions to be linear in the function and its first derivative. Two useful references here are B. Friedman, *Principles and Techniques of Applied Mathematics*, Wiley, New York, 1956, pp. 172–174, and E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw Hill, New York, 1955, pp. 188–193. The proposer remarks that for higher order ordinary differential equations, nothing seems to be known about the maximum order of the poles of the Green's function. In fact, aside from this problem, the literature seems to stop at second-order poles. The proposer thanks Joseph B. Keller and Samuel N. Karp for valuable discussions, and a referee for suggesting improvements in the problem's formulation.

Solved also by Clark Givens.

Square Clusters of Primitive Lattice Points

6568 [1988, 264]. *Proposed by Itshak Borosh and Douglas Hensley, Texas A & M University*.

A lattice point $(s, t) \in N \times N$ is called primitive if the greatest common divisor of s and t is 1. It is known that the density of primitive lattice points is $6/\pi^2 = 0.6079\ldots$ (Hardy and Wright, *Theory of Numbers*, Theorem 459). Show that the density of lattice points $(a, b) \in N \times N$ for which each of the eight neighbors $(a \pm 1, b), (a, b \pm 1), (a \pm 1, b \pm 1)$ is primitive is

$$\frac{1}{4}\prod(1 - 8/p^2) = 0.012\ldots,$$

where the product is taken over all odd primes p.

Composite solution by the University of South Alabama Problem Group, Mobile, and the proposers. Let us call a lattice point $(a, b) \in N \times N$ having the property

and

$$\Delta_2(t,\lambda) = \left[1 + \cos\lambda\pi - \frac{\lambda\pi}{15}\sin\lambda\pi\right]\left[-23\lambda^{-1}\sin\lambda(\pi - t)\right] + \left[\lambda^{-1}\sin\lambda(\pi - t) + \frac{\pi}{15}\cos\lambda(\pi - t)\right][22 + 23\cos\lambda\pi].$$

Since $\Delta_1(0) = 0$ we have $N(0) = 0$. Furthermore, as $\lambda \to 0$ the first term of $N(\lambda)$ is

$$K_1\lambda + O(\lambda^2)$$

where $K_1 \neq 0$ is *independent* of x, while the second term of $N(\lambda)$ is

$$K_2\lambda + O(\lambda^2),$$

where

$$K_2 = 2x[23(t - \pi)] + 45x\left[\pi - t + \frac{\pi}{15}\right] = (t + 2\pi)x.$$

Hence the zero here is of first order, and the result follows.

Editorial comment. The proposer evidently intended the underlying space to be an interval and the boundary conditions to be linear in the function and its first derivative. Two useful references here are B. Friedman, *Principles and Techniques of Applied Mathematics*, Wiley, New York, 1956, pp. 172–174, and E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw Hill, New York, 1955, pp. 188–193. The proposer remarks that for higher order ordinary differential equations, nothing seems to be known about the maximum order of the poles of the Green's function. In fact, aside from this problem, the literature seems to stop at second-order poles. The proposer thanks Joseph B. Keller and Samuel N. Karp for valuable discussions, and a referee for suggesting improvements in the problem's formulation.

Solved also by Clark Givens.

Square Clusters of Primitive Lattice Points

6568 [1988, 264]. *Proposed by Itshak Borosh and Douglas Hensley, Texas A & M University*.

A lattice point $(s, t) \in N \times N$ is called primitive if the greatest common divisor of s and t is 1. It is known that the density of primitive lattice points is $6/\pi^2 = 0.6079\ldots$ (Hardy and Wright, *Theory of Numbers*, Theorem 459). Show that the density of lattice points $(a, b) \in N \times N$ for which each of the eight neighbors $(a \pm 1, b), (a, b \pm 1), (a \pm 1, b \pm 1)$ is primitive is

$$\frac{1}{4}\prod(1 - 8/p^2) = 0.012\ldots,$$

where the product is taken over all odd primes p.

Composite solution by the University of South Alabama Problem Group, Mobile, and the proposers. Let us call a lattice point $(a, b) \in N \times N$ having the property

Combining (2), (3), (4), (5), and (6), we obtain (for $x > P_R$)

$$A(x) = \frac{x^2}{4} \prod_{p>2} \left(1 - \frac{8}{p^2}\right) + O(xP_R) + O\left(\frac{1}{R \log R}\right). \tag{7}$$

If we now take $R = \frac{1}{2} \log x$, then $x = e^{2R} > P_R$ as required, since by a familiar elementary estimate

$$\log P_R < R \log 4 < \frac{7}{5} R.$$

Since $xP_R < xe^{7R/5} = x^{17/10}$, our result (1) follows from (7).

In a similar way we may obtain an analogous result (with $x^2/4$ replaced by $\pi x^2/16$) for the number of nearprim points $(a, b) \in N \times N$ such that $a^2 + b^2 \leqslant x^2$.

Solved also by Benjamin Max and by Matthias Winter (Germany).

Rings and Functors

6595 [1989, 264]. *Proposed by P. A. Griffith and J. J. Rotman, University of Illinois at Urbana-Champaign.*

An R-module E is *injective* if it is a direct summand of any module containing it. Every module can be imbedded in an injective module. The *injective envelope* $E(M)$ is an injective module containing M such that there is no injective module E' with $M \subset E' \subset E(M)$ and $E' \neq E(M)$. The injective envelope $E(M)$ exists and is unique to isomorphism.

(i) A functor T is called additive if, for every pair of R-modules A and B, we have $T(f + g) = T(f) + T(g)$ whenever $f, g \in \mathrm{Hom}_R(A, B)$. Show that there is a ring R for which there is no additive functor T: R-mod $\to$ R-mod with $T(M) = E(M)$.

(ii) Show that there is a ring R for which there is no functor T: R-mod $\to$ R-mod with $T(M) = E(M)$.

(iii)* Find all rings R for which there exists a functor with the property mentioned in (ii).

Solution by Fred Richman, Florida Atlantic University, Boca Raton. We show that the ring $\mathbb{Z}$ of integers satisfies (i) and (ii); modules over $\mathbb{Z}$ are just abelian groups. Let $\mathbb{Z}_p$ denote the cyclic group of order p. Let $\mathbb{Z}_{p^\infty}$ be the p-primary component of $\mathbb{Q}/\mathbb{Z}$, and observe that $E(\mathbb{Z}_p) = \mathbb{Z}_{p^\infty}$. It suffices to show that there is no functor T: $\mathbb{Z}$-mod $\to$ $\mathbb{Z}$-mod with $T(\mathbb{Z}_p) \cong \mathbb{Z}_{p^\infty}$.

Clearly such a functor cannot be additive on endomorphisms of $\mathbb{Z}_p$ because the identity map on $\mathbb{Z}_p$ is of additive order p, while the identity map on $\mathbb{Z}_{p^\infty}$ is of infinite additive order. We will show that T *must* be additive on endomorphisms of $\mathbb{Z}_p$, thereby reaching a contradiction.

It suffices to show that T preserves the direct sum diagram for $\mathbb{Z}_p \oplus \mathbb{Z}_p$ ([**1**; Prop. 9.5] or [**2**; Prop. 4, page 193]); that is, if ι_1, ι_2, π_1 and π_2 are the injection

and projection maps for that diagram, then

$$(1)\ T(\pi_1)T(\iota_1) = T(\pi_2)T(\iota_2) = 1,$$
$$(2)\ T(\pi_2)T(\iota_1) = T(\pi_1)T(\iota_2) = 0,$$
$$(3)\ T(\iota_1)T(\pi_1) + T(\iota_2)T(\pi_2) = 1.$$

Condition (1) is immediate from the definition of a functor. As $E(0) = 0$, it follows that $T(f) = 0$ if $f = 0$, so condition (2) holds. Now let $\alpha = T(\iota_1)T(\pi_1) + T(\iota_2)T(\pi_2) - 1$. Clearly $\alpha T(\iota_i) = 0$, so, to prove (3), we need only show that the images of the maps $T(\iota_i)$ generate $T(\mathbb{Z}_p \oplus \mathbb{Z}_p)$. But this is true because those images are disjoint copies of $\mathbb{Z}_{p^\infty}$ in $\mathbb{Z}_{p^\infty} \oplus \mathbb{Z}_{p^\infty}$.

The same proof works for the ring of integers modulo 4, with $\mathbb{Z}_p$ replaced by $\mathbb{Z}_2$, and $\mathbb{Z}_{p^\infty}$ by $\mathbb{Z}_4$; this is the smallest ring that satisfies (i) and (ii).

As a partial response to (iii), we remark that for the polynomial ring $R = k[X]$, where k is a perfect field, an additive functor T with $T(M) = E(M)$ can be constructed.

If we rephrase the original question slightly, we can get a positive result that eliminates pathological examples. A true functorial injective envelope would include a natural embedding of A in $T(A)$. We want a superfunctor T of the identity so that if $f\colon A \to B$, then $T(f)\colon T(A) \to T(B)$ extends f. But the following theorem rules out nontrivial cases of that.

THEOREM. *If T is a superfunctor of the identity on the category of R-modules, and $T(A)$ is an injective envelope of A for each R-module A, then every R-module is injective, and T is the identity.*

Proof. If ι is the inclusion of A in $E(A)$, and $f\colon E(A) \to E(B)$ is such that $f\iota = 0$, then $T(f\iota) = T(f)T(\iota) = 0$. But $T(\iota)$ is an isomorphism because it is one-to-one on A, and maps one injective envelope of A into another. Thus $T(f) = 0$, so $f = 0$, and we have shown that zero is the only map from $E(A)$ to $E(B)$ that extends the zero map from A to B. With $B = E(A)/A$, the natural map from $E(A)$ onto B extends the zero map from A to B, hence is zero, so $B = 0$ and A is injective. Thus every R-module is injective.

REFERENCES

1. P. J. Hilton, and U. Stammbach, A Course in Homological Algebra, Springer-Verlag, 1970.
2. S. Mac Lane, Categories for the Working Mathematician, Springer-Verlag, 1971.

Solved also (to the same extent) by R. J. Chapman and the proposers.

The Additive Group Generated by the Rational Unit Vectors

6603 [1989, 455]. *Proposed by Carl Pomerance, University of Georgia, Athens.*

Let G_n denote the set of points in Euclidean n-space which can be reached from the origin in a finite number of steps of unit length, where each step lands on a point with rational coordinates. In other words G_n is the additive subgroup of Q^n generated by those points in Q^n at distance one from the origin.

(a) Show that $G_n = Q^n$ if and only if $n \geqslant 5$.

(b) Obtain an arithmetic characterization of G_1, G_2, G_3, G_4.

and projection maps for that diagram, then

$$(1)\ T(\pi_1)T(\iota_1) = T(\pi_2)T(\iota_2) = 1,$$
$$(2)\ T(\pi_2)T(\iota_1) = T(\pi_1)T(\iota_2) = 0,$$
$$(3)\ T(\iota_1)T(\pi_1) + T(\iota_2)T(\pi_2) = 1.$$

Condition (1) is immediate from the definition of a functor. As $E(0) = 0$, it follows that $T(f) = 0$ if $f = 0$, so condition (2) holds. Now let $\alpha = T(\iota_1)T(\pi_1) + T(\iota_2)T(\pi_2) - 1$. Clearly $\alpha T(\iota_i) = 0$, so, to prove (3), we need only show that the images of the maps $T(\iota_i)$ generate $T(\mathbb{Z}_p \oplus \mathbb{Z}_p)$. But this is true because those images are disjoint copies of $\mathbb{Z}_{p^\infty}$ in $\mathbb{Z}_{p^\infty} \oplus \mathbb{Z}_{p^\infty}$.

The same proof works for the ring of integers modulo 4, with $\mathbb{Z}_p$ replaced by $\mathbb{Z}_2$, and $\mathbb{Z}_{p^\infty}$ by $\mathbb{Z}_4$; this is the smallest ring that satisfies (i) and (ii).

As a partial response to (iii), we remark that for the polynomial ring $R = k[X]$, where k is a perfect field, an additive functor T with $T(M) = E(M)$ can be constructed.

If we rephrase the original question slightly, we can get a positive result that eliminates pathological examples. A true functorial injective envelope would include a natural embedding of A in $T(A)$. We want a superfunctor T of the identity so that if $f\colon A \to B$, then $T(f)\colon T(A) \to T(B)$ extends f. But the following theorem rules out nontrivial cases of that.

THEOREM. *If T is a superfunctor of the identity on the category of R-modules, and $T(A)$ is an injective envelope of A for each R-module A, then every R-module is injective, and T is the identity.*

Proof. If ι is the inclusion of A in $E(A)$, and $f\colon E(A) \to E(B)$ is such that $f\iota = 0$, then $T(f\iota) = T(f)T(\iota) = 0$. But $T(\iota)$ is an isomorphism because it is one-to-one on A, and maps one injective envelope of A into another. Thus $T(f) = 0$, so $f = 0$, and we have shown that zero is the only map from $E(A)$ to $E(B)$ that extends the zero map from A to B. With $B = E(A)/A$, the natural map from $E(A)$ onto B extends the zero map from A to B, hence is zero, so $B = 0$ and A is injective. Thus every R-module is injective.

REFERENCES

1. P. J. Hilton, and U. Stammbach, A Course in Homological Algebra, Springer-Verlag, 1970.
2. S. Mac Lane, Categories for the Working Mathematician, Springer-Verlag, 1971.

Solved also (to the same extent) by R. J. Chapman and the proposers.

The Additive Group Generated by the Rational Unit Vectors

6603 [1989, 455]. *Proposed by Carl Pomerance, University of Georgia, Athens.*

Let G_n denote the set of points in Euclidean n-space which can be reached from the origin in a finite number of steps of unit length, where each step lands on a point with rational coordinates. In other words G_n is the additive subgroup of Q^n generated by those points in Q^n at distance one from the origin.

(a) Show that $G_n = Q^n$ if and only if $n \geqslant 5$.

(b) Obtain an arithmetic characterization of G_1, G_2, G_3, G_4.

Solution by Yoshio Mimura, Kobe University, Kobe, Japan. Let U_n be the set of points in Q^n at distance one from the origin. We must determine the additive group G_n generated by U_n. If $x = (x_1, \ldots, x_n) \in Q^n$ we define the positive integer $\mathrm{dnm}(x)$ to be the least common denominator of $x_1, \ldots, x_n$.

(1) $G_1 = \mathbb{Z}$. This is clear.

(2) $G_2 = \{x \in Q^2 | \mathrm{dnm}(x)$ is a product of primes $\equiv 1 \bmod 4\}$. To see this we take a point $u = (a/t, b/t) \in U_2$ with $t = \mathrm{dnm}(u)$. Then we have $t^2 = a^2 + b^2$ with $\gcd(a, b) = 1$. Hence any prime divisor of t is congruent to $1 \bmod 4$, which implies that G_2 is contained in R_2, the group on the right-hand side. Let t be a product of primes $\equiv 1 \bmod 4$. Then we have $t^2 = a^2 + b^2$ for some integers a and b with $\gcd(a, b) = 1$. Since $(a/t, \pm b/t) \in U_2$ we see by addition that $(2a/t, 0) \in G_2$. Similarly $(2b/t, 0) \in G_2$. Since $\gcd(a, b) = 1$, there exist integers u, v such that $au + bv = 1$. Hence

$$(2/t, 0) = u(2a/t, 0) + v(2b/t, 0) \in G_2.$$

Similarly $(0, 2/t) \in G_2$. We may assume without loss of generality that a is odd and b is even. Since $(a/t, b/t)$, $(2/t, 0)$, and $(0, 2/t)$ all lie in G_2 and

$$(1/t, 0) = (a/t, b/t) - \lfloor a/2 \rfloor (2/t, 0) - \lfloor b/2 \rfloor (0, 2/t),$$

it follows that $(1/t, 0) \in G_2$. Similarly $(0, 1/t) \in G_2$. Since R_2 is generated by the points $(1/t, 0)$ and $(0, 1/t)$, we have $R_2 \subset G_2$.

(3) $G_3 = \{x \in Q^3 | \mathrm{dnm}(x)$ is odd$\}$. Take any point $u = (a/t, b/t, c/t) \in U_3$ with $t = \mathrm{dnm}(u)$. Then $t^2 = a^2 + b^2 + c^2$ with $\gcd(a, b, c) = 1$. Hence t must be odd, which implies that G_3 is contained in R_3, the group on the right-hand side. For the reverse inclusion it is sufficient to show that $(1/t, 0, 0) \in G_3$ for any odd t. Since $t^2 \equiv 1 (\mathrm{mod}\ 8)$, it is known that $t^2 = a^2 + b^2 + c^2$ for some integers a, b and c with $\gcd(a, b, c) = 1$. (See, for example, B. A. Venkov, *Elementary Number Theory*, Wolters-Noordhoff, Groningen, 1970, p. 165 or E. Grosswald, *Representations of Integers as Sums of Squares*, Springer, New York, 1985, p. 52. In fact the number of integers (a, b, c) such that $a^2 + b^2 + c^2 = t^2$ and $\gcd(a, b, c) = 1$ is equal to

$$6t \prod_{p|t} \{1 - (-1)^{(p-1)/2} p^{-1}\},$$

where the product is taken over the primes dividing the odd number t.) Since $(a/t, \pm b/t, \pm c/t) \in U_3$, we see by addition that $(2a/t, 0, 0) \in G_3$. Similarly $(2b/t, 0, 0) \in G_3$ and $(2c/t, 0, 0) \in G_3$. Since $\gcd(a, b, c) = 1$, there exist integers u, v, w with $au + bv + cw = 1$. Hence

$$(2/t, 0, 0) = u(2a/t, 0, 0) + v(2b/t, 0, 0) + w(2c/t, 0, 0) \in G_3.$$

Similarly $(0, 2/t, 0) \in G_3$ and $(0, 0, 2/t) \in G_3$. We may assume without loss of generality that a is odd while b and c are even. Since $(a/t, b/t, c/t)$, $(2/t, 0, 0)$, $(0, 2/t, 0)$ and $(0, 0, 2/t)$ are all in G_3, it follows as in (2) that $(1/t, 0, 0) \in G_3$. Since similarly $(0, 1/t, 0) \in G_3$ and $(0, 0, 1/t) \in G_3$, we have $R_3 \subset G_3$.

(4) $G_4 = \{g, g + e | g \in G_4^0\}$, where $G_4^0 = \{g \in Q^4 | \mathrm{dnm}(g)$ is odd$\}$ and $e = (1/2, 1/2, 1/2, 1/2)$. Take any $u = (a/t, b/t, c/t, d/t) \in U_4$ with $t = \mathrm{dnm}(u)$. Then we have $t^2 = a^2 + b^2 + c^2 + d^2$. If t is odd, then $u \in G_4^0$. If t is even, then t is divisible by 2 but not by 4, and a, b, c, and d are odd. Hence we have $u = v + e$ with $v \in G_4^0$. Therefore $G_4 \subset \{g, g + e | g \in G_4^0\}$. For the reverse inclu-

sion it is enough to show that $G_4^0 \subset G_4$. If t is odd then $4t^2 - 1 \equiv 3 \bmod 8$. Hence, we can write $4t^2 - 1 = a^2 + b^2 + c^2$ for some integers a, b, c. Thus $(1/t, 0, 0, 0) \in G_4$, because $(1/(2t), \pm a/(2t), \pm b/(2t), \pm c/(2t)) \in U_4$. Hence, $G_4^0 \subset G_4$.

(5) $G_n = Q^n$ if $n \geqslant 5$. It is enough to show that $(1/t, 0, \ldots, 0) \in G_n$ for all positive integers t. We can write $4t^2 - 1 = a^2 + b^2 + c^2 + d^2$ with integers a, b, c, d. Then we have $(1/(2t), \pm a/(2t), \pm b/(2t), \pm c/(2t), \pm d/(2t), 0, \ldots 0) \in U_n$. Hence, $(1/t, 0, \ldots, 0) \in G_n$.

Editorial comment. The proposer remarked that the assertion of part (a) is proved in a paper by K. B. Chilakamarri, Unit-distance graphs in rational n-space, *Discrete Math.*, 69 (1988) 213–218

Solved also by R. J. Chapman (England), F. J. Flanigan, Stephen M. Gagola, Jr., and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Japanese Temple Geometry Problems. By Hidetosi Fukagawa and Dan Pedoe. The Charles Babbage Research Centre, P.O. Box 272, St. Norbert Postal Station, Winnipeg, Canada, R3V 1L6, 1989. xvi + 206 pp.

DAN SOKOLOWSKY
Professor Emeritus, Mathematics, Antioch College, Yellow Springs, OH 45387

The first time I ever saw a Japanese temple geometry problem was some years ago in Roger A. Johnson's *Advanced Euclidean Geometry*. In reference to a certain theorem the author mentions that it is of Oriental origin, and that

> It was the custom of Japanese mathematicians to inscribe their discoveries on tablets which were hung in the temples, to the glory of the gods and the honor of the authors.

The custom struck me as a lovely way to display respect for a beautiful idea and share it publicly. Moreover, the theorem itself, I felt, truly justified such regard. Readers can judge for themselves. It stated:

> Let a convex polygon, inscribed in a circle, be divided into triangles by diagonals from one vertex. Then the sum of the radii of the circles inscribed in these triangles is the same, whichever vertex is chosen.

For a long time thereafter I found no other mention of such temple theorems, and assumed few existed with little known about them. So my interest naturally perked up when, in 1984, the Canadian journal *Crux Mathematicorum* began to publish such theorems from time to time as problems for solution. Readers like myself would learn that a particular problem was a temple problem when its solution appeared, since it would be accompanied by pertinent historical comments about the problem, such as the location and date of the tablet on which it had been found, or perhaps a book of the period in which its solution had been discussed. These comments were supplied by the proposer of these problems, Hidetosi Fukagawa, a Japanese scholar who has been devoting himself for the past fifteen years to collecting them and studying their history.

The deeper Fukagawa got into this study the more he felt that this body of work deserved wider attention, and eventually he wrote about it to a number of mathematicians in the West. Fortunately one of them was Dan Pedoe, who, after studying some of the material, had the acuity to appreciate its significance, while also recognizing that much of it was unknown in the West. Knowing that its then editor, the late Léo Sauvé, as well as many of its readers were keen lovers of geometry, he suggested to Fukagawa that he submit some of these theorems as

marvelling at the sheer diversity of these problems, and the imagination and ingenuity of their creators. I recall becoming intrigued years ago by, for example, the well-known Ancient Theorem of Pappus, yet here one finds page after page of such problems, the variations seem endless.

From the pedagogical point of view, while not a textbook, the book provides very desirable collateral reading for an appropriate course, say in geometry or the history of mathematics, besides being a sourcebook for innumerable interesting problems.

It is certainly a most valuable addition to our body of classical geometry and its history.

Symmetries of Culture: Theory and Practice of Plane Pattern Analysis. By Dorothy K. Washburn and Donald W. Crowe, University of Washington Press, Seattle, 1988, x + 301 pp.

DORIS SCHATTSCHNEIDER
Department of Mathematics, Moravian College, Bethlehem, PA 18018

Classification is a tricky business. Yet it is difficult to think of any area of human enterprise in which this is not a fundamental activity. Classification is the way we sort and order, a means to identify and understand. How we classify reflects our biases and priorities—even the most "objectively scientific" modes of classification are rooted in history and tradition, and not immune from cultural and professional bias.

Over ten years ago, a colleague and I curated an art exhibit at the Allentown Art Museum in which we selected a variety of objects from their permanent collection. Fabrics, lace, quilts, decorated jars, silver serving pieces, tiles, plates, trivets, and lithographs were not classified by historical period, media, type of object, region, artist, or any other standard method of display used in art exhibits. Instead, the objects were displayed in groups according to the symmetry displayed in their design. Although it seemed perfectly natural to a mathematician to choose such a classification, it was highly unnatural to those on the museum staff. The mission of that art exhibit was similar to the mission of the book under review.

The collaborating authors Dorothy Washburn, anthropologist, and Donald Crowe, mathematician, undertake a double task: first, convince those in the discipline of anthropology of the usefulness and viability of a relatively new method of classification of decorated objects, and then, teach them how to use the method. The method is one familiar to mathematicians: use symmetry groups to classify the decorative patterns that adorn the objects of study. The analysis of decorative designs in the plane according to their symmetries is a fairly standard exercise in mathematics courses that address the topics of geometric transformations or groups. Although recently a few anthropologists have advanced this method of analysis, its use is by no means standard in the profession.

It is not surprising, then, that the authors spend the largest part of the first chapter ("History and Theory of Plane Pattern Analysis") in making a convincing case for the use of symmetry analysis as an anthropological tool. The points that are made are especially interesting for mathematicians to read and keep in mind as

marvelling at the sheer diversity of these problems, and the imagination and ingenuity of their creators. I recall becoming intrigued years ago by, for example, the well-known Ancient Theorem of Pappus, yet here one finds page after page of such problems, the variations seem endless.

From the pedagogical point of view, while not a textbook, the book provides very desirable collateral reading for an appropriate course, say in geometry or the history of mathematics, besides being a sourcebook for innumerable interesting problems.

It is certainly a most valuable addition to our body of classical geometry and its history.

Symmetries of Culture: Theory and Practice of Plane Pattern Analysis. By Dorothy K. Washburn and Donald W. Crowe, University of Washington Press, Seattle, 1988, x + 301 pp.

DORIS SCHATTSCHNEIDER
Department of Mathematics, Moravian College, Bethlehem, PA 18018

Classification is a tricky business. Yet it is difficult to think of any area of human enterprise in which this is not a fundamental activity. Classification is the way we sort and order, a means to identify and understand. How we classify reflects our biases and priorities—even the most "objectively scientific" modes of classification are rooted in history and tradition, and not immune from cultural and professional bias.

Over ten years ago, a colleague and I curated an art exhibit at the Allentown Art Museum in which we selected a variety of objects from their permanent collection. Fabrics, lace, quilts, decorated jars, silver serving pieces, tiles, plates, trivets, and lithographs were not classified by historical period, media, type of object, region, artist, or any other standard method of display used in art exhibits. Instead, the objects were displayed in groups according to the symmetry displayed in their design. Although it seemed perfectly natural to a mathematician to choose such a classification, it was highly unnatural to those on the museum staff. The mission of that art exhibit was similar to the mission of the book under review.

The collaborating authors Dorothy Washburn, anthropologist, and Donald Crowe, mathematician, undertake a double task: first, convince those in the discipline of anthropology of the usefulness and viability of a relatively new method of classification of decorated objects, and then, teach them how to use the method. The method is one familiar to mathematicians: use symmetry groups to classify the decorative patterns that adorn the objects of study. The analysis of decorative designs in the plane according to their symmetries is a fairly standard exercise in mathematics courses that address the topics of geometric transformations or groups. Although recently a few anthropologists have advanced this method of analysis, its use is by no means standard in the profession.

It is not surprising, then, that the authors spend the largest part of the first chapter ("History and Theory of Plane Pattern Analysis") in making a convincing case for the use of symmetry analysis as an anthropological tool. The points that are made are especially interesting for mathematicians to read and keep in mind as

they teach symmetry and groups in an abstract setting. "Practically every culture in the world is known to decorate at least some portion of its material culture with repeated patterns." [p. 29] Symmetry is certainly one of the most obvious characteristics of decorative art; less obvious perhaps is that it plays a strong role in perception. It is a salient feature of style: the arrangement of design units is at least as important as the individual elements. "Only a limited number of symmetries are appropriate in any given culture, and adherence to these structures is necessary for approval and use." [p. 28] Symmetry classification of motif arrangement provides the anthropologist a systematic measure of "design structure" and provides a tool to organize the information in a way that is replicable; this description can be used to present cultural preferences. "Without explicitly defined classes of phenomena, the formulation of general theories about design relevant to all cultures cannot be achieved." [pp. 40–41]

Decorative art often shows a motif that repeats in two colors; the arrangement of the colors as well as the arrangement of the repeated motif can be important in design analysis. Although crystallographers and mathematicians in the late nineteenth and early twentieth centuries established the technique of categorizing repeating patterns in two (and three) dimensions according to symmetry groups and enumerated all the possibilities, the coloring of such patterns did not become part of their theory until the mid-1950s. In light of the audience to which *Symmetries of Culture* is directed, there is a bit of irony in the fact that the first systematic symmetry analysis and complete enumeration of two-color periodic types of plane-filling mosaics (complete with decorative illustrations) was published in 1935–1936 by the textile physicist H. J. Woods, in *The Journal of the Textile Institute* (Manchester). These "counterchange" patterns are characterized by the property that each symmetry of the underlying uncolored pattern either leaves the colored pattern unchanged (i.e., the symmetry makes black motifs fall on black and white fall on white) or uniformly interchanges the colors (i.e., the symmetry makes all white motifs fall on black and vice versa). The careful presentation of the classification theory of such two-color designs sets this book apart from other symmetry books produced in recent years.

The book is the coffee-table variety—large size, superbly designed and lavishly produced, with printed endpapers and the two-color mosaics of H. J. Woods embossed in gold on the cover. Over 500 illustrations (most of them photographs) give visual instruction in analysis of plane patterns and provide convincing evidence of the wealth of decorative design to be found in cultures around the world, from ancient times to modern. The authors note that the book was ten years in the making—no doubt a large part of that time was spent in gathering the impressive collection of examples of decorative art to illustrate the complete list of one and two-color symmetry groups of repetitive plane patterns. In only a few instances did they fail to find real examples—photos or drawings that show designs painted, printed, woven, incised, or inlaid on actual objects; in those few cases, an "invented design" is provided to illustrate a symmetry group.

Chapters 2 through 6 are devoted to a careful presentation of mathematical terminology (e.g., the four isometries, as well as the restricted meaning (within the book) of such words as "symmetry of a figure," "design," "pattern," "dimension," and "symmetry consistent with color"), the notation for symmetry groups, and detailed instructions on how to carry out the symmetry analysis of a design.

Chapter 4 treats one-dimensional patterns (friezes), of which there are 7 one-color and 17 two-color types; chapter 5 treats two-dimensional patterns (wallpaper patterns), of which there are 17 one-color and 46 two-color types, and chapter 6 treats finite designs (rosettes), of which there are 2 infinite classes of one-color and 3 infinite classes of two-color designs. The authors' highly visual approach is to permeate the text with illustrative step-by-step examples of symmetry analysis of designs and to provide schematic diagrams, tables, yes-no flowcharts to guide identification, and (best of all) photos and drawings of real examples to illustrate each pattern type. This is a book in which the reader can learn general principles just by studying the pictures and diagrams and turn to the neighboring text for details.

Borrowing a classification scheme from one discipline to apply to another can be both illuminating and risky. Illuminating, because it enables one to see with new eyes what is obscured by the tunnel vision of tradition formed within a single discipline. Risky, because the borrowed scheme may impose its own limitations inherent in its assumptions as to what is important or "correct." Washburn and Crowe give convincing reasons for the illumination to be gained by the addition of symmetry analysis to the standard repertoire of anthropological tools. Certain conventions must be made in order to use that tool, however. For example, one must decide when to treat a pattern as finite, or one-dimensional, or two-dimensional; decide when to ignore details or irregularities; decide when a colored motif is to be treated as part of the background and not part of the design. The authors give their reasonable conventions for each of these decisions, and in a final chapter, provide a few caveats. The most important point (in the reviewer's opinion) is saved for the last page—the information in a design that is disregarded for the sake of symmetry analysis may be the most telling for anthropological study. Purposeful irregularities in repetition or in the coloring of motifs may be mandated by the culture or contain encoded information. Indeed, in earlier chapters, the reader must become frustrated at the obvious inadequacy of classification by color symmetry groups to recognize highly orderly colorings of patterns that fail to have the coloring compatible with most or even any symmetries of the pattern. As such, they are classified as having little or no color symmetry. The need for new classification schemes that recognize orderly arrangements and orderly coloring of motifs in a repeated pattern has been noted by Grünbaum and Shephard [**3**], and an attempt to define such a scheme especially suited to fabric design has recently appeared [**2**].

It has often been noted that one of the most formidable aspects of mathematics to "outsiders" is the mysterious notation that is used for identification of objects. Even mathematical insiders deplore poor notation, or unnecessary notation. Unfortunately, the notation situation in the literature on color groups is, in a word, a mess. It reflects the newness of the subject as well as the difficulty in finding a good characterizing notation. Crowe and Washburn do a real service by providing a cross-reference table of five notations in use for color symmetry groups in the plane. The notation they choose to use throughout the book (that of crystallographers Belov and Tarkhova), however, is unfortunate from the point of view of teaching and understanding the algebraic structure of the symmetry groups. (From the point of view of the target audience, the choice of notation may seem inconsequential—by following the yes-no flowcharts, one can assign the correct

classification symbol to a pattern. It is merely a label; it need not be deciphered or understood.) A much more informed notation (with respect to algebraic structure) for color symmetry groups was implicit in a 1961 paper of B. L. van der Waerden and J. J. Burckhardt and later developed by M. Senechal [**4**]. Given a periodic pattern colored with k colors so that all elements of the symmetry group G of the uncolored pattern are compatible with the coloring, one identifies a subgroup H of G that leaves invariant the set of all motifs of a single chosen color. The index of H in G is k; the Senechal notation for the color symmetry group of the pattern is $[G : H] = k$. When $k = 2$, H is necessarily normal in G; for this case, H. S. M. Coxeter introduced the quotient group notation G/H. For all but one of the 46 two-color groups, the pair G, H completely determines the color symmetry group. Crowe and Washburn discuss Coxeter's notation, but do not use it; yet it is a notation that gives far more information on the arrangement of colors and is easy to use. (First, view a pattern as uncolored and determine its symmetry group G, then squint and focus only on the motifs of one color and determine the symmetry group H of that one-color pattern.)

Why is this book being reviewed here? It was written for archaeologists, anthropologists, and art historians, but the authors have taken care in their presentation of the geometry of symmetry and color symmetry analysis. Aside from the fact that mathematicians and their students need to be more aware of unconventional applications of mathematics, it is a book that is a wonderful resource. It is also a natural choice for a text to teach an introductory course on one- and two-color plane symmetry groups. The level and purpose of such a course will determine how much mathematical detail will need to be supplemented. The book includes proofs of only two theorems: there are only four isometries of the Euclidean plane and there are exactly 7 classes of one-dimensional periodic patterns. The book concludes with an extensive bibliography.

REFERENCES

1. L. Baker and D. Schattschneider, The Perceptive Eye: Art and Mathematics, Allentown Art Museum, Allentown, 1979.
2. B. Grünbaum, Periodic ornamentation of the fabric plane: lessons from Peruvian fabrics, *Symmetry* 1(1990) 45–68.
3. B. Grünbaum and G. Shephard, Tilings and Patterns, W. H. Freeman, New York, 1987.
4. M. Senechal, Point groups and color symmetry, *Zeit. Kristall.* 142 (1975) 1–23.
5. B. L. van der Waerden and J. J. Burckhardt, Farbgruppen, *Zeit. Kristall.*, 115 (1961) 231–234.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P*, L.** *Visualization in Teaching and Learning Mathematics.* Eds: Walter Zimmermann, Steve Cunningham. MAA Notes Ser., No. 19. MAA, 1991, viii + 224 pp, (P). [ISBN: 0-88385-071-0] A superb collection of profusely illustrated papers on various aspects of visualization in mathematics—its role in intuition and insights; its risks and benefits in proofs; its potential in the curriculum. Examples from calculus, linear algebra, differential equations, analysis, differential geometry, complex analysis, numerical analysis, and stochastic processes. A visual and intellectual delight that will surely stimulate imaginative pedagogical innovation. LAS

General, P. *Las Matematicas En Costa Rica.* Ed: Angel Ruiz Zuñiga. Departamento de Publicaciones Universidad Nacional (Escuela de Matematica, Universidad de Costa Rica, San Jose, CR), 1990, x + 533 pp, (P). Proceedings of the October 1990 third Costa Rican national mathematics conference, including sections on pure mathematics, applied mathematics, mathematics of computation, mathematics education, history of mathematics, and philosophy of mathematics. A supplementary volume contains various late papers. LAS

General, P*. *Actions for Renewing U.S. Mathematical Sciences Departments.* National Research Council, 1990, vii + 29 pp, (P). Advice to chairs of university mathematics departments about steps they can take to "become active participants" in the effort to renew U.S. mathematics. Focuses on setting goals, educational initiatives (at all levels), research support (both disciplinary and interdisciplinary), and strategies to alter the reward system. This report, sponsored by the (research) Board on Mathematical Sciences at the National Research Council, is part of the wide-ranging effort of the National Academy of Sciences to improve mathematics education at all levels. LAS

Mathematics Appreciation, S*(13-14), L*. *More Mathematical Morsels.* Ross Honsberger. Dolciani Math. Expos., No. 10. MAA, 1991, xii + 322 pp, $16 (P). [ISBN: 0-88385-313-2] The sixth of Honsberger's delightful Dolciani volumes: expansive solutions for intriguing elementary problems selected from *Crux Mathematicorum* (a journal of undergraduate problem solving published by the Canadian Mathematical Society) together with "gleanings" from Murray Klamkin's "Olympiad Corner" which appeared regularly in that journal. A pure delight for problem solvers; a rich resource for math clubs. LAS

Education, P. *Algebridge: Concept Based Instructional Assessment.* College Board. Janson, 1990, xvii + 414 pp, $44.90 (P). [ISBN: 0-939765-34-9] An innovative set of diagnostic assessment units intended to help prepare students for the transition from arithmetic to algebra. Intended to be used as a supplement to any standard course or text; consists of a rich teacher resource guide plus sets of student assessment booklets dealing with such topics as fractions, negatives, proportional reasoning, variables, inequality, and operations. Prepared as part of the College Board's Educational EQuality Project intended to widen the access of secondary school students to college. LAS

History, S(16-17), P, L. *History of Continued Fractions and Padé Approximants.* Claude Brezinski. Springer-Verlag, 1991, 551 pp, $79. [ISBN: 0-387-15286-5] A history up to 1939. The exposition consists of a more-or-less chronological reporting of developments interspersed with passages of original texts and papers. Of value to any teacher of elementary number theory. The 150-page bibliography must surely contain every relevant reference. SG

History, S*, P, L*. *Huygens and Barrow, Newton and Hooke.* V.I. Arnol'd. Transl: Eric J.F. Primrose. Birkhauser Boston, 1990, 118 pp, (P). [ISBN: 0-8176-2383-3] A penetrating historical probe that reveals contemporary perspectives on deep mathematical problems explored by Newton and his contemporaries: evolutes and quasicrystals; *Principia* and Kirkwood gaps; Kepler's second law and Abelian integrals. Written in a rich expository style; full of historical detail enriched by contemporary mathematical wisdom. LAS

Linear Algebra, T(16-17: 1). *Fundamentals of Matrix Computations.* David S. Watkins. Wiley, 1991, xiii + 449 pp, $51.95. [ISBN: 0-471-61414-9] Designed to be used as a textbook for courses in numerical linear algebra, it can also serve as a reference for those desiring an introduction to a particular topic or technique. It covers error analysis, elimination schemes, least-squares problems, eigenvalues and eigenvectors, and the singular value decomposition. AO

Linear Algebra, T(14-15: 1). *Algèbre Linéaire.* Joseph Grifone. Cepadues-Editions, 1990, viii + 439 pp, (P). [ISBN: 285-428-239-6] A rather inclusive introduction to the theory of vector spaces, usually finite-dimensional, over an arbitrary field. Exercises, hints on solutions. JD-B

Group Theory, T(18: 1), S, P. *Quasigroups and Loops: Theory and Applications.* Eds: O. Chein, H.O. Pflugfelder, J.D.H. Smith. Sigma Ser. in Pure Math., V. 8. Heldermann Verlag, 1990, xii + 568 pp, DM 148. [ISBN: 3-88538-008-0] A collaborative effort with fourteen chapters, all by different authors, sharing a common belief that non-associativity has "now become susceptible" to systematic analysis. Wide range of topics, both algebraic and geometric. Quite contemporary. Index, extensive bibliography. JS

Group Theory, T(16-17: 1), S, L. *Quasigroups and Loops: Introduction.* Hala O. Pflugfelder. Sigma Ser. in Pure Math., V. 7. Heldermann Verlag, 1990, 147 pp, DM 68. [ISBN: 3-88538-007-2] Assuming only an elementary course in abstract algebra, this text provides a basic introduction to quasigroups and loops. Chapters on relations to geometry, isotopy, Moufang loops. Exercises, bibliography, index. JS

Algebra, P. *Lie Algebras and Related Topics.* Eds: Georgia Benkart, J. Marshall Osborn. Contemp. Math., V. 110. AMS, 1990, xxxvi + 313 pp, $44 (P). [ISBN: 0-8218-5119-5] Proceedings of a research conference held May 22-June 1, 1988 at the University of Wisconsin at Madison. LC

Algebra, T(17-18: 1), S, L. *Elements of Mathematics: Algebra II, Chapters 4-7.* N. Bourbaki. Transl: P.M. Cohn, J. Howie. Springer-Verlag, 1990, vii + 461 pp, $98. [ISBN: 0-387-19375-8] Translation from the French of the chapters on polynomials and rational functions, commutative fields, ordered groups and fields, and modules over principal ideal domains. Manages to preserve the spirit, style, notation, and mannerisms of the original. Exercises, historical notes, bibliography, index. JS

Calculus, T(13: 1). *Calculus, Volume 1* (Instructor's Preliminary Edition). Thomas P. Dick, Charles M. Patton. PWS-Kent, 1991, xvi + 366 pp, (P). [ISBN: 0-534-92779-3] The first semester's part of the new calculus text being prepared by the NSF-funded Oregon State University Calculus Curriculum Project. Stresses intelligent use of technology (to avoid "becoming a victim of its pitfalls"); multiple representations of functions (graphical, numerical, symbolic); visualization (a machine-generated graph "as the first step instead of the last one"); and mathematical modelling. Typical syllabus, elaborated in an inviting, "lean" style; makes only metaphorical (e.g., "zoom") reference to calculators or computers. A standard supplement (excerpted in this edition) on using technology in calculus discusses important issues (e.g., roundoff, viewing windows) independent of hardware and provides additional technology-dependent examples and exercises. LAS

Calculus, S(13-14: 1-3). *Uses of Technology in the Mathematics Curriculum.* Benny Evans, Jerry Johnson. Oklahoma State University, 1990, 188 pp, (P). Outlines of over fifty laboratory projects for use with computer algebra systems to supplement calculus, linear algebra, and differential equations. Written without reference to specific software; opens with a section providing mini-reviews of available software. Based on two NSF-sponsored workshops at Oklahoma State University. LAS

Calculus, T(13: 2). *Single Variable Calculus, Second Edition.* James Stewart. Brooks/Cole, 1991, xvii + 736 pp, $47.25. [ISBN: 0-534-14532-9] A concise presentation of traditional topics with early coverage of trigonometric functions; introduces topics with intuitive descriptions when appropriate; incorporates Pólya-like problem solving strategies. This edition (a revision of *Calculus,* TR, November 1987) reorganizes treatment of limits, has earlier applications of integration, introduces exponential functions before logarithmic, and includes more numerical integration. JNC

Partial Differential Equations, P, L. *Lecture Notes in Mathematics-1450: Functional-Analytic Methods for Partial Differential Equations.* Eds: H. Fujita, T. Ikebe, S.T. Kuroda. Springer-Verlag, 1990, vii + 251 pp, $29 (P). [ISBN: 0-387-53393-1] Proceedings of a conference honoring Tosio Kato held at the University of Tokyo, July 1989. Contains 17 papers and a list of recent publications of Kato. MLR

Operator Theory, P. *Nonlinear Superposition Operators.* Jürgen Appell, Petr P. Zabrejko. Tracts in Math., V. 95. Cambridge University Pr, 1990, 311 pp, $59.50. [ISBN: 0-521-36102-8] Gives definition of superposition operators. Analyzes the boundedness, compactness, and differentiability of a function and the related properties of the operator determined by the function. Involves nonlinear analyses on ideal spaces, Lebesgue spaces, Sobolev spaces, and other functional spaces. LS

Operator Theory, P. *Proceedings Seminar 1986-1987: Mathematical Structures in Field Theories.* G.G.A. Bäuerle, *et al.* CWI Syllabus, V. 26. Mathematisch Centrum, 1990, iii + 169 pp, Dfl. 48 (P). [ISBN: 90-6196-387-7] Six papers from lectures given at the Centre for Mathematics and Computer Science in Amsterdam ranging from Dirac's formalism to supersymmetric field theory. SP

Operator Theory, P. *Operator Theory: Proceedings of the 1988 GPOTS-Wabash Conference.* Eds: J.B. Conway, B.B. Morrel. Pitman Res. Notes in Math. Ser., V. 225. Longman Sci. & Technical (US

Distr: Wiley), 1990, 183 pp, $34 (P). [ISBN: 0-582-06190-3] Proceedings of a May 1988 conference in Indianapolis focusing on operator theory. Includes papers by Arveson, Ghatage, Ji and S. Pedersen, G.K. Pedersen, Peligrad, Price, and Teleman. JS

Analysis, T(14-17: 1, 2), L. *An Introduction to Abstract Analysis.* W.A. Light. Chapman & Hall, 1990, xiii + 194 pp, $29.95 (P); $65. [ISBN: 0-412-31090-2; 0-412-31080-5] Assumes only an advanced calculus background. Gently introduces the reader to abstract analysis with an eye toward setting the stage for further readings in functional analysis. Contains many examples and exercises and, at first viewing, appears quite readable. KS

Algebraic Geometry, P. *Helices and Vector Bundles: Seminaire Rudakov.* A.N. Rudakov, *et al.* Transl: A.D. King, P. Kobak, A. Maciocia. London Math. Soc. Lect. Note Ser., V. 148. Cambridge University Pr, 1990, 143 pp, $24.95 (P). [ISBN: 0-521-38811-2] A collection of twelve seminar talks delivered in Moscow on the use of helices to study exceptional vector bundles on an n-dimensional variety. SG

Algebraic Geometry, P. *Spectral Theory and Analytic Geometry Over Non-Archimedean Fields.* Vladimir G. Berkovich. Math. Surveys & Monographs, No. 33. AMS, 1990, ix + 169 pp, $53. [ISBN: 0-8218-1534-2] Exposition of the author's work on non-Archimedean analytic spaces (extending the concept of a rigid analytic space), and the basics of the spectral theory of bounded linear operators on such spaces. SG

Algebraic Geometry, P. *G-Functions and Geometry.* Yves André. Aspects of Math., V. E13. Friedr. Vieweg & Sohn (US Distr: GLP International Vieweg), 1989, xii + 229 pp, DM 48 (P). [ISBN: 3-528-06317-3] An exposition of Siegel's theory of G-functions and its applications to diophantine approximation and arithmetic algebraic geometry. SG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1439: Automorphism Groups of Compact Bordered Klein Surfaces, A Combinatorial Approach.* Emilio Bujalance, *et al.* Springer-Verlag, 1990, xiii + 201 pp, $22 (P). [ISBN: 0-387-52941-1] The authors consider the question of realizing a finite group of a certain type (abelian, nilpotent, solvable) as the automorphism group of a compact Klein surface. The book is well-written, contains an extensive bibliography, and is relatively self-contained. SG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1447: Cohomology of Arithmetic Groups and Automorphic Forms.* Eds: J.-P. Labesse, J. Schwermer. Springer-Verlag, 1990, 356 pp, $39 (P). [ISBN: 0-387-53422-9] A collection of fourteen articles, all presented at a 1989 conference. Includes a survey article by Schwermer, as well as papers by Wallach, Ash and Borel, Harder, Oda, and others. SG

Algebraic Geometry, P. *Algebraic Geometry.* Eds: H. Kurke, J.H.M. Steenbrink. Kluwer Academic, 1990, viii + 305 pp, $99. [ISBN: 0-7923-0934-0] A special issue of *Compositio* containing fifteen papers covering a wide range of topics in algebraic geometry. SG

Differential Geometry, P. *Geometric Analysis and Computer Graphics.* Eds: P. Concus, R. Finn, D.A. Hoffman. Math. Sci. Res. Inst. Public., V. 17. Springer-Verlag, 1991, ix + 203 pp, $39.80. [ISBN: 0-387-97402-4] Twenty-two papers from a workshop on differential geometry, calculus of variations, and computer graphics at the Mathematical Sciences Research Institute in Berkeley, May 23-25, 1988. Includes many computer-generated figures, including eight pages of color plates. DFA

Operations Research, T(15-17), L. *Operations Research: Applications and Algorithms, Second Edition.* Wayne L. Winston. PWS-Kent, 1991, xvi + 1262 pp. [ISBN: 0-534-92495-6] *Second Edition* of an excellent text for a first course in operations research. Full of examples that motivate mathematical solutions and theory. Changes include: 3D geometry of LPs feasible region, matrix generators of LPs, graphical approach to sensitivity analysis, rewritten transshipment and shortest path problems, new section on Karmarkar's method, network simplex method, cutting plane method for integer programming, method of feasible directions for nonlinear programming problems, Bayesian inference with the normal distribution, extended coverage of inventory theory, analytical hierarchy principle, first passage times in Markov chains, "open" queueing networks, and a chapter on forecasting. MK

Optimization, T(17-18). *Existence and Optimality of Competitive Equilibria.* Charalambos D. Aliprantis, Donald J. Brown, Owen Burkinshaw. Springer-Verlag, 1990, xii + 284 pp, $35 (P). [ISBN: 0-387-52866-0] A "systematic exposition ... on general equilibrium models with an infinite number of commodities." Discusses existence and optimality of competitive equilibria in Walrasian general equilibrium model with finite number of households, firms, and commodities, and extends to the infinite commodities case. Uses Riesz spaces to model the infinite case; includes chapter on basic Riesz theory. Contains a number of exercises after each section. MK

Optimization, T(17: 1, 2), L. *Foundations of Integer Programming.* Harvey M. Salkin, Kamlesh Mathur. North-Holland (US Distr: Elsevier Science), 1989, xix + 755 pp, $55. [ISBN: 0-444-01231-1] An introductory survey of integer programming techniques and applications. Covers cutting plane techniques, enumerative methods, group theoretic techniques, and specialized algorithms for the knapsack, set covering, fixed charge, and traveling salesman problems. The bibliography contains more than 600 entries. AO

Optimization, P. *Large-Scale Numerical Optimization.* Eds: Thomas F. Coleman, Yuying Li. SIAM, 1990, 255 pp, $38.50 (P). [ISBN: 0-89871-268-8] Fifteen papers from a workshop held at Cornell University on October 19-20, 1989. The papers are organized into four sections: applications, linear and quadratic problems, sparse and nonlinear problems,

and parallel approaches. AO

Probability, P. *Mazimum Entropy and Bayesian Methods.* Ed: Paul F. Fougère. Fund. Theories of Physics, V. 39. Kluwer Academic, 1990, xiii + 480 pp, $129. [ISBN: 0-7923-0928-6] Text from proceedings of Ninth Annual Maximum Entropy Workshop held at Dartmouth College, August 14-18, 1989. Thirty-two essays include nice discussion of ontological vs. epistemological probability (Edwin Jaynes), Bayesian methods applied to astrophysical problems, discussions of model selection, parameter estimation, construction of priors, supernova analysis, drug absorption, etc. MK

Stochastic Processes, P. *Counting Process Systems: Identification and Stochastic Realization.* P.J.C. Spreij. CWI Tract, V. 71. Mathematisch Centrum, 1990, 135 pp, Dfl. 38.50 (P). [ISBN: 90-6196-388-5] A slightly revised version of the author's thesis, this monograph presents results on recursive parameter estimation algorithms and realization problems for counting process systems. AO

Stochastic Processes, P. *Adaptive Algorithms and Stochastic Approximations.* Albert Benveniste, Michel Métivier, Pierre Priouret. Transl: Stephen S. Wilson. Appl. of Math., V. 22. Springer-Verlag, 1990, xi + 365 pp, $59. [ISBN: 0-387-52894-6] An introduction to the field of adaptive algorithms—mathematical techniques which repeatedly apply a recurrence formula to converge on the correct numerical result. The text introduces the basic ideas—convergence rates, parameter tracking, validation—and demonstrates these ideas with a number of examples drawn from different areas of mathematics. GMS

Elementary Statistics, T(15-17: 2), L. *Biostatistics: A Foundation for Analysis in the Health Sciences, Fifth Edition.* Wayne W. Daniel. Wiley, 1991, xii + 740 pp, $49.95. [ISBN: 0-471-52514-6] *Fifth Edition* of a first-principles statistics text (*First Edition,* TR, February 1975). Aimed at advanced undergraduates, first-year graduate students, and health-care professionals. Covers standard, frequentist, statistical methodology; probability distributions, point and interval estimation, testing, one and two-way ANOVA, simple and linear regression, and nonparametric statistics. Additions to this edition include exploratory data analysis; new exercises; free diskettes with large data sets; incorporation of Minitab, complete with appendix of commands and syntax. Includes examples with detailed, well-exposed problems. MK

Elementary Statistics, T*(13-14: 1, 2), L. *Introduction to Statistics: The Nonparametric Way.* Gottfried E. Noether. Texts in Stat. Springer-Verlag, 1991, xii + 414 pp, $49.50. [ISBN: 0-387-97284-6] Successor to the author's 1976 text *Introduction to Statistics: A Nonparametric Approach* (TR, August-September 1976). Provides an alternative to the standard pre-calculus statistics course. Includes the classical normal-theory procedures after each topic has been more thoroughly discussed using a nonparametric approach. Includes a Minitab section at the end of all but two of the twenty chapters. RSK

Computational Statistics, S, C. *Learning Data Analysis with Data Desk.* Paul F. Velleman. WH Freeman, 1989, xii + 353 pp, $26.95 (P). [ISBN: 0-7167-2072-8] A simple graphics-oriented Macintosh statistics package that provides multiple windows displaying various summary statistics, graphs (e.g., histograms, scatterplots), random samples, inference, ANOVA, and regression. Tools permit entry and editing of data (the student version does not allow importing of data except from the full professional version), extensive editing of the graphical displays, and saving of both text and graphical output for use in other programs. "HyperViews" (submenus) make options visible; "hot" data makes the consequences of changes instantly visible in various windows. A valuable exploratory tool for standard elementary topics. LAS

Computational Statistics, T(17-18), S, P, L. *Generalized Additive Models.* T.J. Hastie, R.J. Tibshirani. Mono. on Stat. & Appl. Prob., V. 43. Chapman & Hall, 1990, xv + 335 pp, $55. [ISBN: 0-412-34390-8] Discusses intuition, application, and theory for various nonparametric regression-like models. Covers a multiplicity of smoothing techniques with exposition of trade-offs, choice of smoothing parameters, cross-validation, and special topics such as nonlinear smoothers, Kriging, and resistant smoothing. Additional chapters cover generalized additive models, transformations (in particular, ACE), additive models for matched case-control data, proportional-hazards, seasonal time-series, model selection, and adaptive regression splines. MK

Statistics, T(16: 1), L. *A First Course in the Theory of Linear Statistical Models.* Raymond H. Myers, Janet S. Milton. Ser. in Stat. & Decision Sci. PWS-Kent, 1991, x + 342 pp. [ISBN: 0-534-91645-7] Fine introduction to the theory of estimation and testing in full rank and less-than-full rank linear models. Preliminary chapters review matrix algebra and present quadratic forms and their distributions. Includes illustrative SAS output at the end of the chapters. RWJ

Statistics, T(13: 1, 2), L. *Statistics for the Life Sciences.* Myra L. Samuels. Dellen, 1989, xi + 597 pp. [ISBN: 0-02-405501-8] Standard introduction to descriptive and inferential statistics assuming an elementary algebra background. Covers correlation, regression, one-way ANOVA, and basic nonparametric tests. Also considers questions of experimental design and simultaneous inference. Contains a wealth of real data sets. RWJ

Statistics, T(16-17), L. *Design and Analysis of Experiments, Third Edition.* Douglas C. Montgomery. Wiley, 1991, xvii + 649 pp, $56.95. [ISBN: 0-471-52000-4] Standard topics for first course in experimental design: completely randomized, randomized block, Latin square, incomplete, factorial, fractional factorial, nested and split-plot designs. Also, regression analysis, response surface methods,

and analysis of covariance are discussed. This edition reorganizes material on 2^k and fractional factorial designs. Discusses Taguchi approach to parameter design. Numerous exercises and terse examples. (*Second Edition,* TR, August-September 1984.) MK

Statistics, P. *Statistical Analysis of Measurement Error Models and Applications.* Eds: Philip J. Brown, Wayne A. Fuller. Contemp. Math., V. 112. AMS, 1990, xii + 248 pp, $53 (P). [ISBN: 0-8218-5117-9] Collection of papers presented at a conference held June 10-16, 1989 at Humboldt State University. Considers linear and nonlinear measurement error models; a few papers deal with computational aspects and robustness questions. Includes a short survey paper of work in this area. RWJ

Statistics, P. *Lecture Notes in Statistics-65: Non-Standard Rank Tests.* Arnold Janssen, David M. Mason. Springer-Verlag, 1990, 252 pp, $34 (P). [ISBN: 0-387-97484-9] Theoretical monograph presenting "a methodology to construct rank tests for models where the standard regularity assumptions do not hold and to study their asymptotic properties." RSK

Statistics, P. *Lecture Notes in Statistics-64: Minimax Solutions in Sampling from Finite Populations.* Siegfried Gabler. Springer-Verlag, 1990, v + 132 pp, $19 (P). [ISBN: 0-387-97358-3] Minimax and modified minimax solutions for survey sampling problems with various types of parameter spaces. RWJ

Statistics, P. *Aspects of Risk Theory.* Jan Grandell. Ser. in Stat. Springer-Verlag, 1991, x + 175 pp, $39. [ISBN: 0-387-97368-0] Generalizes the classical stochastic model of an insurance business by allowing the occurrence of claims to be described by point processes which are not necessarily Poisson. Intended for actuaries with a good knowledge of classical risk theory, and for probabilists with an interest in modern ruin theory. RWJ

Statistics, P. *Statistics in Science: The Foundations of Statistical Methods in Biology, Physics and Economics.* Eds: Roger Cooke, Domenico Costantini. Boston Stud. in Philo. of Sci., V. 122. Kluwer Academic, 1990, ix + 184 pp, $80. [ISBN: 0-7923-0797-6] Proceedings of an international conference held in Luino, Italy. Based on "the conviction that all non-deductive inferences in science are ultimately statistical inferences." Contains ten papers illustrating probabilistic reasoning in a variety of fields. Note price! RSK

Statistics, S(18), P. *State Space Modeling of Time Series, Second, Revised and Enlarged Edition.* Masanao Aoki. Springer-Verlag, 1990, xvii + 323 pp, $29.50 (P). [ISBN: 0-387-52870-9] ARMA models are converted into state space forms, properties of state space models are discussed, and two alternate decompositions of Hankel matrices are used in constructing estimators. The new material in this edition include an expanded description of the statistical properties, especially asymptotic properties, of the proposed estimators. RWJ

Statistics, T(14: 1), L. *Applied Regression Analysis for Business and Economics.* Terry E. Dielman. Ser. in Stat. & Decision Sci. PWS-Kent, 1991, ix + 573 pp. [ISBN: 0-534-92238-4] Introduces simple and multiple linear regression models (including one and two-factor ANOVA) assuming a prior course in statistics and college algebra. Discusses the checking of model assumptions, variable selection, the use of indicator variables, and applications to time series data. Data sets, some of which are revisited throughout the text, are analyzed with the aid of SAS and Minitab. RWJ

Statistics, T(18: 2), P. *Introduction to Statistical Pattern Recognition, Second Edition.* Keinosuke Fukunaga. Computer Sci. & Scient. Computing. Academic Pr, 1990, xiii + 591 pp, $59.95. [ISBN: 0-12-269851-7] Presents statistical decision making and estimation, both parametric and nonparametric, from the viewpoint of pattern recognition. Discusses procedures for classifier design, parameter estimation, and density estimation in detail. Concludes with a chapter on parametric and nonparametric clustering. RSK

Statistics, T(18: 1), P*. *Generalized Multivariate Analysis.* Fang Kai-Tai, Zhang Yao-Ting. Springer-Verlag, 1990, xi + 220 pp, $59. [ISBN: 0-387-17651-9] Summarizes the theory developed during the past fifteen years of multivariate analysis based on elliptically contoured distributions, which includes classical (normal) multivariate analysis as a special case. Good set of references. RSK

Statistics, P, L. *Linear Models for Multivariate, Time Series, and Spatial Data.* Ronald Christensen. Texts in Stat. Springer-Verlag, 1991, xii + 317 pp, $49.50. [ISBN: 0-387-97413-X] Theory of multivariate linear models with a variety of applications. Examples from published, real-world, referenced sources. Due to breadth of topics covered, discussion of each is necessarily brief. Topics range from discrimination, principal components and factor analysis to time series, both frequency and time domain, and Kriging for spatial data. MK

Statistics, S(17-18), P, L. *Analysis of Repeated Measures.* M.J. Crowder, D.J. Hand. Mono. on Stat. & Appl. Prob., V. 41. Chapman & Hall, 1990, x + 257 pp, $55. [ISBN: 0-412-31830-X] Details wide variety of methods for analyzing data involving repeated measures. Includes analysis of longitudinal and spatial studies. Discusses multivariate techniques, regression models, hierarchical models, crossover experiments, categorical data models, and more. Contains examples complete with data and graphs, but there are no exercises. Makes use of BMDP and SPSS^x, including source code, output, and analysis. MK

Programming, P, L. *Wicked Problems, Righteous Solutions: A Catalogue of Modern Software Engineering Paradigms.* Peter DeGrace, Leslie Hulet Stahl. Computing Ser. Prentice-Hall, 1990, ix + 244 pp, (P). [ISBN: 0-13-590126-X] A monograph recounting and commenting on the various models of software development, includes the sequen-

tially phased "waterfall" model, which is most commonly used; variations on the waterfall model; prototyping; "all-at-once" problem solving. The authors' stimulating remarks question contemporary approaches. RB

Programming, T(15), S. *Object-Oriented Modeling and Design.* James Rumbaugh, *et al.* Prentice-Hall, 1991, xii + 500 pp. [ISBN: 0-13-629841-9] An introductory text on the topic of software development using the concept of object-oriented software design. In this technique real world elements are modelled using simulated software objects, which are then manipulated using the techniques of message passing, class structure, and inheritance. It is claimed that this is a superior technique for constructing efficient, correct, and maintainable software. GMS

Languages, P. *Encapsulated PostScript: Application Guide for the Macintosh and PCs.* Peter Vollenweider. Prentice-Hall, 1990, xvii + 226 pp, (P). [ISBN: 0-13-275843-1] An eclectic collection of vendor information, user hints, and programming examples intended to introduce the new standard PostScript page description language EPS. Useful details obscured by poor organization; the lengthy index may be of some help. LAS

Languages, P, L*. *Inside Smalltalk, Volume II.* Wilf R. LaLonde, John R. Pugh. Prentice-Hall, 1991, xvii + 553 pp. [ISBN: 0-13-465964-3] The second of two volumes providing an in-depth introduction to Smalltalk-80, an object-oriented programming environment. This volume focuses on the construction of interactive graphical interfaces. It examines the model-view-controller paradigm, window management, and mouse interaction in the Smalltalk-80 system. AO

Languages, P, L. *Object Oriented Programming with Smalltalk/V.* Duško Savić. Ser. in Comput. & Their Applic. Ellis Horwood (US Distr: Simon & Schuster), 1990, 340 pp. [ISBN: 0-13-040692-9] An introduction to the principles of object-oriented programming using the Smalltalk/V language and environment. Language features and programming techniques are introduced in the context of examples from a variety of applications areas. AO

Algorithms, P. *Parallel Algorithms for Matrix Computations.* K.A. Gallivan, *et al.* SIAM, 1990, ix + 197 pp, $14.50 (P). [ISBN: 0-89871-260-2] "Three papers that collect, describe, or reference an extensive selection of important parallel algorithms for matrix computations The first paper contains a general perspective on modern parallel and vector architectures and the way in which they influence algorithm design. The second paper is primarily concerned with parallel algorithms for solving symmetric positive definite sparse linear systems. The final paper consists of an extensive bibliography on parallel and vector numerical algorithms." LC

Computer Systems, P, L. *MH & zmh: E-mail for Users and Programmers.* Jerry D. Peek. O'Reilly & Associates, 1990, xxviii + 553 pp, $27.95 (P). [ISBN: 0-937175-63-3] The *MH* Message Handling System is a collection of public domain electronic mail programs for computers running UNIX. *zmh* is an X Window System interface to *MH*. This book is a comprehensive guide to the features of *MH*. AO

Computer Systems, P. *TUGboat: The Communications of the TEX Users Group.* Ed: Barbara Beeton, Christina Thiele. Conf. Proc., V. 10, No. 4. TEX Users Group (PO Box 9506, Providence, RI 02940), 1989, 315 pp, $75 (P). Proceedings of the tenth annual meeting of the TEX Users Group held at Stanford University, August 20-23, 1989. Includes 30 papers describing projects using METAFONT, TEX, LATEX, and related software. AO

Computer Graphics, S(16-18), P. *3D Interactive Computer Graphics: The Hierarchical Modelling System HIRASP.* Wim J. Teunissen, Jan van den Bos. Ser. in Comput. & Their Applic. Ellis Horwood, 1990, ix + 229 pp. [ISBN: 0-13-921867-X] Complete presentation of the HIRASP software system, which is entirely oriented to raster graphics hardware and device-independent in that realm. HIRASP defines 3D objects for rendering instead of drawing primitives, treats domain and color at the same level, contains a complete 3D viewing pipeline, provides a set of graphics input facilities, and allows interactive drawing manipulation. DFA

Artificial Intelligence, P. *Advances in Artificial Intelligence: Natural Language and Knowledge-based Systems.* Ed: Martin Charles Golumbic. Springer-Verlag, 1990, xiv + 303 pp, $39. [ISBN: 0-387-97355-9] Fourteen papers selected from recent artificial intelligence conferences in Israel, *viz.*, the Israeli National Conferences on Artificial Intelligence (1987, 1988), and the applications track of the Bar-Ilan Symposium on the Foundations of Artificial Intelligence. Topics include design and architecture, morphology and linguistic analysis, intelligent tutoring and training systems, semantics for constraint-based systems. DFA

Artificial Intelligence, P, L. *Verifying and Validating Personal Computer-Based Expert Systems.* A. Terry Bahill. Prentice-Hall, 1991, xi + 205 pp, (P). [ISBN: 0-13-957457-3] Topics: extracting knowledge from the system, verifying the knowledge base, validating the system, testing the system. Includes a study of the "pregnant man" problem, sensitivity analyses, use of neural networks. References, glossary, collection of thought problems and possible solutions, animal classification expert system. Good for self-study; chapters are independent of each other. DFA

Artificial Intelligence, S(18), P. *Intelligent Systems: State of the Art and Future Directions.* Zbigniew W. Ras, Maria Zemankova. Ser. in AI. Ellis Horwood, 1990, vii + 529 pp. [ISBN: 0-13-465931-7] Twenty papers on knowledge representation, inference mechanisms, reasoning under uncertainty, machine learning, expert systems, intelligent information systems, the role of artificial intelligence in computer applications, and the impact and future of in-

telligent systems. Some are surveys, some are new developments. DFA

Computer Science, T(14-15: 1, 2). *Data Structures and Program Design in C.* Robert L. Kruse, Bruce P. Leung, Clovis L. Tondo. Prentice-Hall, 1991, xv + 525 pp. [ISBN: 0-13-725649-3] C version of the first author's text using Pascal. Suitable for both the ACM CS2 and CS7 (Data Structures and Algorithm Analysis) courses. Assumes some experience with C, but provides a succinct introduction. Includes many examples of substantial length, exercises, and programming projects. Emphasizes problem specification; program design, analysis, testing, verification, and correctness; major principles of software engineering; data abstraction. Good discussion of recursion and its uses. DFA

Computer Science, T(15-16). *Parallel Programs for the Transputer.* Ronald S. Cok. Prentice-Hall, 1991, xii + 242 pp, (P). [ISBN: 0-13-651480-4] Introduces the Transputer Architecture—a massively parallel, MIMD-based, private memory computer architecture. Also teaches how to program in Occam-2, the parallel language of the transputer. Both of these systems are widely used in academia and industry, and are excellent vehicles for learning the concepts of parallel programming. GMS

Applications, P. *Continuation and Bifurcations: Numerical Techniques and Applications.* Eds: Dirk Roose, Bart De Dier, Alastair Spence. NATO ASI Ser. C, V. 313. Kluwer Academic, 1990, xiii + 426 pp, $132. [ISBN: 0-7923-0855-7] Contains papers from ten countries in Europe and North America. The central goal is to describe bifurcation problems and numerical techniques. The papers extend over spectrum theories, numerical analysis, software, and applications. LS

Applications, P, L. *Mathematica for the Sciences.* Richard E. Crandall. Addison-Wesley, 1991, xiii + 300 pp. [ISBN: 0-201-51001-4] A collection of examples and exploratory exercises from mathematics, physics, biology, chemistry, and engineering that illustrate the use of *Mathematica* as a problem-solving tool in the sciences. The examples explore the symbolic, graphical, and numerical capabilities of *Mathematica.* Programming techniques are developed on an "as needed" basis. AO

Applications, T(17-18: 2), P, L. *Log-Linear Models.* Ronald Christensen. Texts in Stat. Springer-Verlag, 1990, xi + 408 pp, $49. [ISBN: 0-387-97398-2] Bridges gap between Fienberg's monograph and Bishop, Fienberg, and Holland's definitive reference. This excellent text contains many fully-examined examples and exercises along with readable text. Discusses theoretical and computational aspects of the applications. Includes full discussion of logistic linear models and a brief discussion of generalized linear models. MK

Applications (Biological Science), P. *Some Mathematical Questions in Biology: Sex Allocation and Sex Change: Experiments and Models.* Ed: Marc Mangel. Lect. on Math. in the Life Sci., V. 22. AMS, 1990, ix + 205 pp, $41 (P). [ISBN: 0-8218-1172-X] Proceedings of the 1989 symposium held at the annual meeting of the AIBS in Toronto, Canada, August 7, 1989. Topics include sex allocation in plants, male allocation and cost of biparental sex in a parasitic trematode, population genetics of sex allocation through exact population models, and various aspects of hermaphroditism in animals. MK

Applications (Communication Theory), T(17) P. *Discrete Cosine Transform: Algorithms, Advantages Applications.* K.R. Rao, P. Yip. Academic Pr, 1990, xviii + 490 pp, $69.95. [ISBN: 0-12-580203-X] First, basic theory: definitions, general properties, relations to the Karhunen-Loeve transform, fast algorithms, two-dimensional algorithms, performance. Then, the heart of the book: over 200 pages describing 23 major applications. Concludes with 90 pages of computer programs, a list of appropriate VLSI chip manufacturers, and an extensive, up-to-date bibliography. Good exercise sets throughout. Includes many figures, including color plates. DFA

Applications (Physics), T*(14-15), L. *The Principles of Electromagnetic Theory.* Attay Kovetz. Cambridge University Pr, 1990, xvii + 221 pp, $49.50; $19.95 (P). [ISBN: 0-521-39106-7; 0-521-39997-1] The title suggests that this is just one more introductory text in electromagnetic theory. However, the difference here is that the author presents the subject as a classical theory independent of the atomic constitution of matter. He begins with a formal presentation of the laws of *e-m* theory (in tensor form) and then establishes a link between *e-m* theory, classical mechanics, and thermodynamics. The last half of the book deals with applications to electrostatics, dielectrics, etc. The level of the text is appropriate for second or third-year students with a background in multivariable calculus. Not that many problems. MPR

Applications (Social Science), P. *Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs.* Eds: Donald R. Brown, J.E. Keith Smith. Recent Res. in Psychology. Springer-Verlag, 1991, xxvi + 202 pp, $35 (P). [ISBN: 0-387-97451-2] Collection of seven essays honoring Clyde Coombs. Introduction and biography by Amos Tversky. Topics include axiomatic approach to measurement theory; social dilemmas, economic self-interest, and evolutionary theory; probabilistic dimensionality; judgments of relative importance in decision making; validating the dimensional structure of psychological spaces. MK

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, St. Olaf; LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; MK: Michael Kahn, St. Olaf; RSK: Richard S. Kleber, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LS: Li Sheng, Macalester; LAS: Lynn Arthur Steen, St. Olaf.

SUBSCRIBE TO

UME TRENDS

News and reports on Undergraduate Mathematics Education

Keep up with what's happening in Undergraduate Mathematics Education.

UME TRENDS is conducting a subscription drive for its third volume beginning March 1991.

Whether you are receiving it now or not, you must subscribe in order to keep your issues coming.

We must receive a minimum number of subscriptions in order to keep publishing UME TRENDS.

SUBSCRIBE NOW!

Subscriptions are $12 per year for six issues.

Copy or clip the adjoining form and mail it to arrive by February 8, 1991.

UME TRENDS, Volume 3, 1991. Please check the appropriate category and fill in the remaining information.

___ $12 subscriber within the United States ___ $20 subscriber outside the United States

Check Method of Payment
__ Check or money order __ Visa __ MasterCard

Name _____ Card Number _____

Address _____ Card Expiration date _____

_____ Signature _____

Note that subscriptions must be prepaid.

MAIL TO:
AMERICAN MATHEMATICAL SOCIETY
P. O. Box 1571, Annex Station
Providence, Rhode Island 02901-1571

When was the last time a computer program helped you *think* about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Saunders College Publishing Salutes

Mathematics Awareness Week

April 21st–27th

and its sponsors

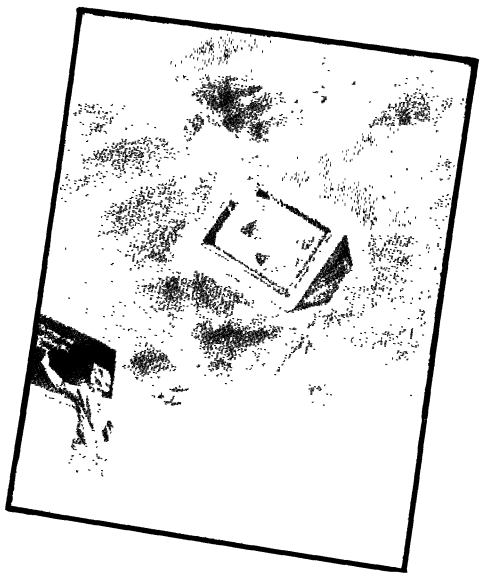
The American Mathematical Society
The Mathematical Association of America
The Society for Industrial and Applied Mathematics

The Powers of Mathematics

SAUNDERS COLLEGE PUBLISHING

The Curtis Center, Independence Square West, Philadelphia, PA 19106

Career Information from the MAA



MATHEMATICAL SCIENTISTS AT WORK

If you have difficulty clarifying career opportunities in the mathematical sciences for your students, then let these 14 practitioners speak for you. This attractive booklet provides career information about statistics, actuarial science, operations research, and computer science as well about mathematics. Each essay is accompanied by a photograph of the author and by additional information relevant to the occupation. The booklet also contains general information about careers in mathematics and sources where additional information may be obtained.

This is what some of our writers have to say to students:

...when seeking a mathematical position, don't be put off the trail by titles of jobs or work units that may not seem mathematical. Mathematicians can pop up, and do mathematical work, all over the place.

Michael Weiss, *Agricultural Economist*

The attempts to build expert systems, which can perform tasks such as medical diagnosis or operating a computer, show the need for mathematical models of how computer systems can store data and "reason" about knowledge.

Maria Klawe, *Computer Scientist*

This valuable booklet should find its way into every mathematics department and guidance office in the country. Be sure that your students get their hands on a copy.

32 pp., 1990, ISBN 0-88385-452-X

Price: 1-9 copies $5.00 each

Price: 10-49 copies $4.50 each

Price 50 or more copies, $4.00 each

Price 100 or more copies, $3.00

Catalog Number MSW

CAREERS IN THE MATHEMATICAL SCIENCES

This pamphlet contains condensations of six of the essays in "Mathematical Scientists at Work" plus sources for additional information. It is ideally suited for distribution to high school or college mathematics classes and to mathematics clubs. Take some with you when giving a talk to student groups and make them available to students who visit your college campus—you will find many uses for these handy pamphlets.

Price: 1-99 copies 25 cents each

Price: 100 or more copies 20 cents each

Catalog Number CMS

Order From



Mathematical Association of America
1529 18th Street, N.W.
Washington, D.C. 20036

NEW! VERSION 2

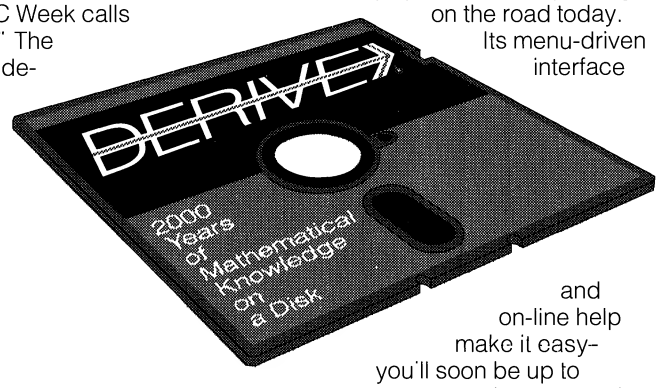
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE
May 29, 1990
Derive, Version 1.6

DERIVE® – A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

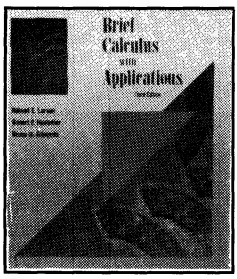
And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

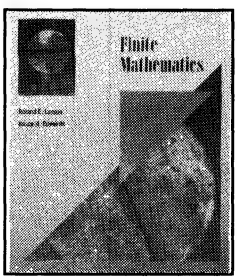
Handcrafted Software for the Mind

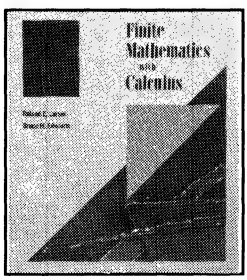
Soft Warehouse INC
HONOLULU · HAWAII

Another Year of Excellence
D. C. Heath Mathematics in 1991









*The popular **Brief Calculus**— now part of a remarkable series!*

Brief Calculus with Applications
Third Edition
1991 Cloth 812 pages
and
Brief Calculus with Applications
Alternate Third Edition
1991 Cloth 648 pages

Roland E. Larson and **Robert P. Hostetler**, both of The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida

Finite Mathematics
1991 Cloth 500 pages
and
Finite Mathematics with Calculus
1991 Cloth 926 pages

Roland E. Larson, The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida

Also new for 1991:

Understandable Statistics
Fourth Edition
Charles Henry Brase, Regis College
Corrinne Pellillo Brase,
Arapahoe Community College
1991 Cloth 704 pages est.

Elementary Linear Algebra
Second Edition
Roland E. Larson, The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida
1991 Cloth 592 pages

Arithmetic for College Students
Sixth Edition
D. Franklin Wright, Cerritos College
1991 Cloth 528 pages

Mathematics for Elementary Teachers:
A Balanced Approach
Second Edition
Eugene F. Krause, University of Michigan
1991 Cloth 928 pages est.

Basic Mathematics
Third Edition
David Novak, Simmons College
1991 Paper 656 pages est.

For more information, call us toll-free:
(800) 235-3565.

D. C. Heath and Company
125 Spring Street
Lexington, MA 02173

Just Published from HBJ!

INTRODUCTORY ALGEBRA
and
INTERMEDIATE ALGEBRA
RONALD HATTON and
GENE R. SELLERS
both of *Sacramento City College*
Each Paperbound, 1991

ELEMENTARY ALGEBRA
Second Edition
and
INTERMEDIATE ALGEBRA
Second Edition
YOSHIKO YAMATO and
MARY JANE CORDÓN
both of *Pasadena City College*
Each Hardcover, 1991

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN,
Drexel University, and
ARNOLD SHAPIRO
Hardcover, 1991

FINITE MATHEMATICS
WILLIAM OWEN and
FRED CUTLIP
both of *Central Washington University*
Hardcover, 1991

When you asked for a sensible, no-frills, mainstream calculus textbook at a reduced cost for your students, we heard you...

CALCULUS
One and Several Variables
ROBERT ELLIS and
DENNY GULICK
both of *University of Maryland*
Hardcover, 1991 $29.95 (Net)!

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of *Vassar College*
Hardcover, 1991

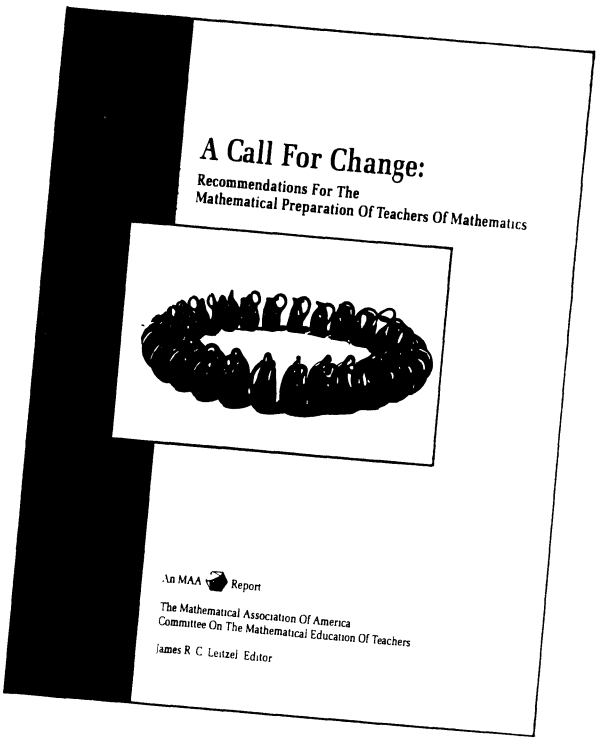
ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.,
Michigan State University
Hardcover, 1991

ABSTRACT ALGEBRA
DENNIS KLETZING,
Stetson University
Hardcover, 1991

HBJ

HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

An MAA Report on Teacher Preparation



A CALL FOR CHANGE
Recommendations for the Mathematical Preparation of Teachers of Mathematics

James R. C. Leitzel, Editor

How can we improve the teaching and learning of mathematics in our schools to better prepare our students for the future? We can begin by making some changes in the way our teachers learn and teach mathematics. A CALL FOR CHANGE, a new MAA Report, offers a set of recommendations that come from a vision of ideal mathematics teachers in classrooms of the 1990's and beyond. It describes the collegiate mathematical experiences that a teacher needs in order to meet this vision.

A CALL FOR CHANGE discusses standards common to the preparation of mathematics teachers at all levels; Elementary School Level (K–4), Middle School Level (5–8), and Secondary School Level (9–12). It shows what type of preparation teachers need to help them:

- (i) view mathematics as a system of interrelated principles,
- (ii) communicate mathematics accurately verbally, and in writing,
- (iii) understand the elements of mathematical modeling,
- (iv) understand and use calculators and computers appropriately in the teaching and learning of mathematics, and
- (v) appreciate the development of mathematics both historically and culturally.

Mathematics continues to be a dynamic changing discipline. There is new mathematics that can be exciting for young people to learn and technology provides new approaches for teachers to engage students in the teaching and learning of mathematics. The mathematical preparation of teachers must adapt to these changing realities. Order your copy now of this important set of guidelines.

60 pp., 1991, ISBN 0-88385-072-9

List: $7.00

Catalog Number CFC

Order From



Mathematical Association of America
1529 18th Street, N.W.
Washington, D.C. 20036

Recreations for the Mathematically Minded...

FROM W. H. FREEMAN AND COMPANY

CAN YOU WIN?

The Real Odds for Casino Gambling, Sports Betting, and Lotteries

MIKE ORKIN

Mike Orkin lays out the hard, cold, and often amusing facts about gambling in this offbeat guide for novice and expert alike. He tells of the odds of winning and possible strategies for roulette, keno, slots, sports betting, blackjack, lotteries, horse racing, and "prisoner's dilemma." Enjoyable reading and an invaluable reference.

March, 181 pages, paper, 0-7167-2155-4, $11.95

MATHEMATICA® IN ACTION

STAN WAGON

Extend the power of Mathematica to the limits of the possible with alternative methods to generate three-dimensional graphics, iterative graphics, and animations. This book contains many valuable shortcuts, complete programs with line-by-line explanations, and hundreds of advanced examples worked in detail.

March, 352 pages, 133 illustrations, 4 pp. color plates, paper 0-7167-2202-X, $29.95; cloth 0-7167-2229-1, $39.95

VISUALIZATION

The Second Computer Revolution

RICHARD MARK FRIEDHOFF

Visualization explains why and how we use the computer to make images. Included in this comprehensive guide for the general reader are sections on computer animation, simulation of complex phenomena such as black holes and snowflakes, medical imaging, procedural graphics, architecture with CAD systems, numerical experiments, and the new "artificial realities."

March, 216 pages, 200 illustrations, including 136 in full color, paper, 0-7167-2231-3, $24.95

KEEP YOUR EYE ON THE BALL

The Science and Folklore of Baseball

ROBERT G. WATTS and A. TERRY BAHILL

Could Sandy Koufax's curve really "fall off a table?" Did Ted Williams actually see the ball hit his bat when he connected? Engineers Watts and Bahill put some of baseball's cherished myths to the test of scientific scrutiny in a highly informative and entertaining guide.

March, 220 pages, 101 illustrations, paper, 0-7167-2248-8, $12.95

FRACTALS

An Animated Discussion with Edward Lorenz and Benoit B. Mandelbrot

A 63-Minute Video

H.-O. PEITGEN, H. JÜRGENS, D. SAUPE, and C. ZAHLTEN

This video presents exciting color animations of the world of fractals, chaos, and self-similarity. An accompanying 28-page booklet provides the algorithms and mathematical equations that underlie the fractals featured in the video.

1990, 63-minute full-color video, VHS format, 0-7167-2213-5, $59.95

THE MAGIC MACHINE

A Handbook of Computer Sorcery

A. K. DEWDNEY

This is the second collection of Dewdney's popular "Scientific Recreations" columns, drawn from Scientific American. The author discusses some of today's hottest topics, including chaos, computer viruses, and artificial landscapes.

1989, 357 pages, 107 illustrations, cloth, 0-7167-2125-2, $23.95; paper, 0-7167-2144-9, $15.95

AVAILABLE AT FINE BOOKSTORES or order directly from: W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84104.

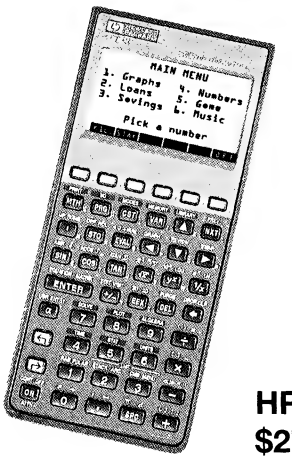
W. H. FREEMAN AND COMPANY

The book publishing arm of Scientific American

INTRODUCING
E.Z. MATH

For use with HP'S HP 48SX

For Teachers, Students, and other Professionals.



HP 48SX
$275.00

Graphs. E.Z. Math covers the entire high school and college graphing curriculum from elementary algebra to advanced calculus. The student can choose from 182 families of equations, inequalities, functions, systems of equations and systems of inequalities, in rectangular, polar and parametric form, all laid out and arranged in an easy-to-use, logically organized system of menus. E.Z. Math makes the graphic analysis of polynomial, rational, trigonometric, hyperbolic, logarithmic and exponential functions quick and easy.

Numbers. E.Z. Math makes it easy to handle numbers of various types. Elementary and high school students will find it easy to compute all the factors and the prime factorization of a natural number. In addition, finding the greatest common factor, least common multiple and average of any set of natural numbers is a snap. Any set of rational numbers, whether in whole number, fraction of mixed number form, may be added, subtracted, multiplied or divided yielding an answer expressed as a fraction or mixed number in lowest terms or as an integer. Any set of complex numbers may be added, subtracted, multiplied, divided, or raised to a power. Any number of terms of the sequences of perfect nth powers, triangular numbers, binomial coefficients, Fibonacci numbers and multiples of a number can be easily computed.

Savings. E.Z. Math makes it easy to solve problems involving a single deposit or repeated deposits to a savings account, certificate of deposit (C.D.), term deposit (T.D.), money market account or other such investment. Anyone interested in saving money can quickly and easily compute the amount accumulated, the amount that must be deposited, the number of years necessary to keep the money on deposit and the annual interest rate.

Loans. E.Z. Math makes it easy to solve problems involving fixed rate mortgage on houses, condos, co-ops, and other real estate investments. Anyone wanting to purchase a home can quickly computed the monthly mortgage payment, the amount that can be borrowed, the number of years needed to pay off the loan and the annual interest rate and see a detailed breakdown of each payment in a complete amortization table.

Recreation. E.Z. Math provides a couple of strategy number games and an introduction to the musical capabilities of the HP48SX. These are provided both for fun and as an encouragement to explore some of the many features of the calculator, including a bit of the built-in rich programming language.

A quote from a review: "The software takes you by the hand and marks your path to all of E.Z. Math's remarkable features."

Order today and take advantage of this marvelous combination of handheld hardware and software.

- HP 48SX....$275.00 each (add $5.00 for shipping and handling)
- E.Z. Math software....$125.00 each (add $3.00 for shipping and handling.)
- E.Z. Math software plus HP 48SX if purchased as a package....$390.00 (includes shipping and handling)
- Orders of 10 or more packages....$375.00 each (includes shipping and handling)

To order write or call: SMI, Dept. AMM, 250 West New Street, Suite 48, Kingsport, TN 37660
Phone: 800-234-0123 or 615-378-4821 FAX: 615-245-8962

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 5 May 1991

Contents

(ISSN 0002-9890)

ARTICLES

The Cauchy Problem for the Wave Equation with Distribution Data: an Elementary Approach CALVIN H. WILCOX 401

Continuous Nowhere-Differentiable Functions—an Application of Contraction Mappings HIDEFUMI KATSUURA 411

LETTERS TO THE EDITOR 417

NOTES

A Note on Finding a Strict Saddlepoint DANIEL BIENSTOCK, FAN CHUNG, MICHAEL FREDMAN, ALEJANDRO A. SCHÄFFER, PETER W. SHOR, AND SUBHASH SURI 418

Circumscribed Circles ROBERT OSSERMAN 419

Stirling's Series and Bernoulli Numbers ELIAS Y. DEEBA AND DENNIS M. RODRIGUEZ 423

THE TEACHING OF MATHEMATICS

A New Scheme for Multiple-Choice Tests in Lower-Division Mathematics BRUCE R. JOHNSON 427

A Simple Proof of the Weierstrass Approximation Theorem J. H. LINDSEY II 429

Statistics Within Departments of Mathematics at Liberal Arts Colleges THOMAS L. MOORE AND JEFFREY A. WITMER 431

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 437

Advanced Problems and Solutions 445

REVIEWS

Representations and Characters of Finite Groups. By M. J. Collins MARK STEVEN MAZUR 453

Theorist. By Prescience Corporation FRANK WATTENBERG 455

TELEGRAPHIC REVIEWS 461

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $32 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

The Cauchy Problem for the Wave Equation with Distribution Data: an Elementary Approach

CALVIN H. WILCOX, *University of Utah, Salt Lake City, UT 84112*



CALVIN H. WILCOX received his Ph.D. at Harvard under Garrett Birkhoff and H. Levine. He has held positions at Caltech, the Universities of Wisconsin, Arizona and Utah, and Denver University, and has served as visiting professor at the Universities of Geneva, Liege, Kyoto, Stuttgart and Bonn. He received an Alexander von Humboldt Senior U.S. Scientist Award in 1976. He has written some 60 research articles on boundary value problems, spectral analysis of differential operators in Hilbert spaces, wave propagation and scattering theory as well as three monographs: Scattering Theory for the d'Alembert Equation in Exterior Domains, Scattering Theory for Diffraction Gratings, and Sound Propagation in Stratified Fluids.

Introduction. The Cauchy problem for the wave equation in two space dimensions asks for a function $u(x, y, t)$ that satisfies

$$u_{tt} - u_{xx} - u_{yy} = h(x, y, t) \qquad \forall\,(x, y, t) \in R^2 \times [t_0, \infty) \tag{1}$$

and

$$u(x, y, t_0) = f(x, y), \qquad u_t(x, y, t_0) = g(x, y) \qquad \forall\,(x, y) \in R^2, \tag{2}$$

where f, g, h are prescribed functions on their respective domains. It was discovered in the early nineteenth century that a formal solution of the problem can be constructed by quadratures. The result is usually written

$$\begin{aligned} u(x, y, t) &= \frac{1}{2\pi} \iint_{D(x,y,t-t_0)} \frac{g(\xi, \eta)}{\sqrt{(t - t_0)^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta \\ &\quad + \frac{\partial}{\partial t} \left\{ \frac{1}{2\pi} \iint_{D(x,y,t-t_0)} \frac{f(\xi, \eta)}{\sqrt{(t - t_0)^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta \right\} \\ &\quad + \int_{t_0}^{t} \left\{ \frac{1}{2\pi} \iint_{D(x,y,t-\tau)} \frac{h(\xi, \eta, \tau)}{\sqrt{(t - \tau)^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta \right\} d\tau, \end{aligned} \tag{3}$$

where $D(x, y, \tau) = \{(\xi, \eta) | (\xi - x)^2 + (\eta - y)^2 \leq \tau^2\}$ is the disk in R^2 with center at (x, y) and radius τ. This relation is usually attributed to M. Parseval (?–1836). It can be derived from the well-known formula of S. D. Poisson (1781–1840) for the solution of the corresponding problem in three space dimensions; see Hadamard [4] for references to the original literature. In older texts, equation (3) is usually called *Parseval's formula*, or the *Poisson-Parseval formula*; see, for example [2], [8].

Classical theory established that the Parseval formula defines a solution of the Cauchy problem whenever the data f, g, h have sufficiently many continuous

derivatives. This theory is too restrictive for the study of phenomena such as shock waves, which correspond to data that are nondifferentiable or even discontinuous. An extended theory which is applicable to these cases became available in 1950–1951 with the advent of L. Schwartz's theory of distributions [7]. This in turn made it possible to study physically important solutions, such as the fundamental solution, whose Cauchy data f, g, h are themselves distributions.

The solvability of the Cauchy problem (1), (2) for Cauchy data f, g, h which are arbitrary distributions has been known for many years; see for example [5, p. 139, Theorem 5.6.3]. Unfortunately, the published expositions of this result that are known to the author stand at the end of long and intricate abstract theories which tend to obscure the simplicity of the final results.

The purpose of this article is to formulate the Cauchy problem (1), (2) for arbitrary distribution data, and to prove the uniqueness and existence of a solution, by an elementary method that does not use concepts or theorems of functional analysis. The method is constructive and shows that Parseval's formula (3), when properly interpreted, yields the solution of the Cauchy problem for all distribution data f, g, h. The exposition uses only the simplest concepts and facts of distribution theory. The necessary material may be found in the first 17 pages of Hörmander's book [5] of 1963. A more leisurely treatment of the same material is given in the first third of the book of Friedlander [1].

Classical Solutions of the Cauchy Problem. A *classical solution* of the Cauchy problem is a function $u(x, y, t) \in C^2(R^2 \times [t_0, \infty))$ that satisfies conditions (1) and (2). Several methods are available to prove the uniqueness of classical solutions. One is the energy method; see, for example, [9], p. 280. A second method is to prove that any classical solution must be given by Parseval's formula (3). Thus one has

THEOREM 1. *The Cauchy problem* (1), (2) *has at most one classical solution.*

The conditions $f \in C^2(R^2)$, $g \in C^1(R^2)$, $h \in C(R^2 \times [t_0, \infty))$ are clearly necessary for the existence of a classical solution. However, it is known that they are not sufficient. Sufficient conditions may be derived from Parseval's formula. The simplest result of this type is

THEOREM 2. *Parseval's formula defines a classical solution* $u(x, y, t)$ *of the Cauchy problem* (1), (2) *for all* $f \in C^3(R^2)$, $g \in C^2(R^2)$, $h \in C^2(R^2 \times [t_0, \infty))$.

Sketch of the Proof. The linearity of the Cauchy problem (1), (2) implies that if $u_{(f,g,h)}$ denotes the solution for data f, g, h then

$$u_{(f,g,h)} = u_{(f,0,0)} + u_{(0,g,0)} + u_{(0,0,h)}. \tag{4}$$

Moreover, if the integral operator $P_{x,y,t}(g)$ is defined by

$$P_{x,y,t}(g) = \frac{1}{2\pi} \iint_{D(x,y,t)} \frac{g(\xi, \eta)}{\sqrt{\left(t^2 - (x-\xi)^2 - (y-\eta)^2\right)}} \, d\xi \, d\eta, \tag{5}$$

then Parseval's formula implies that

$$u_{(0,g,0)}(x, y, t) = P_{x,y,t-t_0}(g), \tag{6}$$

$$u_{(f,0,0)}(x, y, t) = (\partial/\partial t) P_{x,y,t-t_0}(f), \tag{7}$$

and

$$u_{(0,0,h)}(x, y, t) = \int_{t_0}^{t} P_{x,y,t-\tau}(h(\cdot, \cdot, \tau))\, d\tau, \tag{8}$$

where the last integrand denotes the action of the integral operator $P_{x,y,t-\tau}$ on $h(\xi, \eta, \tau)$ with τ fixed. The relations (7), (8) are called Stokes's rule and Duhamel's principle, respectively. The proof of Theorem 2 will now be sketched for the special case that $f = 0$, $h = 0$, $t_0 = 0$ and $g \in C^2(R^2)$. The general case will then follow from the decomposition (4)–(8).

Consider then the function $v(x, y, t)$ which is defined for $(x, y, t) \in R^2 \times [0, \infty)$ by

$$v(x, y, t) = \frac{1}{2\pi} \iint_{D(x,y,t)} \frac{g(\xi, \eta)}{\sqrt{t^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta, \tag{9}$$

where $g \in C^2(R^2)$. In this classical form the integral is not well suited to proving the continuity and differentiability of v. The simple change of variables $\xi \to x + t\xi$, $\eta \to y + t\eta$ yields the more useful representation

$$v(x, y, t) = \frac{t}{2\pi} \iint_{D(0,0,1)} \frac{g(x + t\xi, y + t\eta)}{\sqrt{1 - \xi^2 - \eta^2}}\, d\xi\, d\eta. \tag{10}$$

The main tools in proving that (10) defines a classical solution of the Cauchy problem are Lebesgue's dominated convergence theorem and Fubini's theorem. Note that $1/\sqrt{1 - \xi^2 - \eta^2}$ is unbounded but integrable on the disk $D(0, 0, 1)$. In fact it is in the Lebesgue class $L_p(D(0, 0, 1))$ for $1 \le p < 2$. Thus for (x, y, t) in an ε-neighborhood $N(x_0, y_0, t_0) = \{(x, y, t) | (x - x_0)^2 + (y - y_0)^2 + (t - t_0)^2 \le \varepsilon^2\}$ the integrand in (10) is bounded by $M/\sqrt{1 - \xi^2 - \eta^2} \in L_1(D(0, 0, 1))$ where M is the maximum of $|g(\xi, \eta)|$ on the disk $D(x_0, y_0, t_0 + 2\varepsilon)$. The continuity of v at (x_0, y_0, t_0) follows easily by Lebesgue's dominated convergence theorem.

The existence and continuity of the derivatives of v follow similarly. For example, if we define

$$v_1(x, y, t) = \frac{t}{2\pi} \iint_{D(0,0,1)} \frac{g_x(x + t\xi, y + t\eta)}{\sqrt{1 - \xi^2 - \eta^2}}\, d\xi\, d\eta, \tag{11}$$

then $v_1 \in C(R^2 \times [0, \infty))$ and Fubini's theorem implies that

$$v(x, y, t) - v(x_0, y, t) = \int_{x_0}^{x} v_1(\xi, y, t)\, d\xi, \tag{12}$$

whence $v_1 = v_x \in C(R^2 \times [0, \infty))$. The same method can be used to complete the proof that $v \in C^2(R^2 \times [0, \infty))$. Finally, the verification that v satisfies (1) follows from a careful application of the divergence theorem to the integral representation for $v_{tt} - v_{xx} - v_{yy}$. Details are left to the reader.

The method outlined above can also be used to prove

COROLLARY 3. *The Cauchy problem* (1), (2) *has a classical solution* $u \in C^\infty(R^2 \times [t_0, \infty))$ *if and only if* $f, g \in C^\infty(R^2)$ *and* $h \in C^\infty(R^2 \times [t_0, \infty))$.

Distribution Solutions of the Cauchy Problem. A distribution-theoretic formulation of the Cauchy problem can be found by beginning with a classical solution $u(x, y, t)$, and calculating $u_{tt} - u_{xx} - u_{yy}$ in the distributional sense. To this end it is natural to extend the definitions of u and h from $R^2 \times [t_0, \infty)$ to all of R^3 by defining $u(x, y, t) \equiv h(x, y, t) \equiv 0$ $\forall (x, y, t) \in R^2 \times (-\infty, t_0)$. With this understanding, one may regard u and h as elements of $\mathscr{D}'(R^3)$ = the set of distributions on R^3 and f, g as elements of $\mathscr{D}'(R^2)$. Then, if $\phi \in \mathscr{D}(R^3) = C_0^\infty(R^3)$ = the set of Schwartz testing functions, then the basic rules of distributional calculus and integration by parts give

$$\begin{aligned}(u_{tt} - u_{xx} - u_{yy})(\phi) &= u(\phi_{tt} - \phi_{xx} - \phi_{yy}) \\ &= \int_{t_0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, y, t)(\phi_{tt} - \phi_{xx} - \phi_{yy})\, dx\, dy\, dt \\ &= \int_{t_0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y, t)\phi(x, y, t)\, dx\, dy\, dt \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{-f(x, y)\phi_t(x, y, t_0) \\ &\qquad + g(x, y)\phi(x, y, t_0)\}\, dx\, dy \\ &= h(\phi) + (f \otimes \delta'_{t_0})(\phi) + (g \otimes \delta_{t_0})(\phi), \end{aligned} \tag{13}$$

where $f \otimes \delta'_{t_0}$, $g \otimes \delta_{t_0}$ are the tensor products of distributions $f, g \in \mathscr{D}'(R^2)$ with distributions $\delta_{t_0}, \delta'_{t_0} \in \mathscr{D}'(R)$; see, for example, [1]. Thus the classical solution u, when regarded as a distribution on R^3, satisfies

$$u_{tt} - u_{xx} - u_{yy} = k \text{ in } \mathscr{D}'(R^3), \tag{14}$$

and

$$\operatorname{supp} u \subset R^2 \times [t_0, \infty), \tag{15}$$

where

$$k = h + f \otimes \delta'_{t_0} + g \otimes \delta_{t_0} \in \mathscr{D}'(R^3), \tag{16}$$

and

$$\operatorname{supp} k \subset R^2 \times [t_0, \infty). \tag{17}$$

Note that conditions (14), (15) are meaningful for any distributions u and k in $\mathscr{D}'(R^3)$. This suggests the

DEFINITION. *A distribution solution* of the Cauchy problem (1), (2) with distribution data $f, g \in \mathscr{D}'(R^2)$ and $h \in \mathscr{D}'(R^3)$ with $\operatorname{supp} h \subset R^2 \times [t_0, \infty)$ is a distribution $u \in \mathscr{D}'(R^3)$ that satisfies (14), (15) and (16).

Note that in the distributional formulation of the Cauchy problem there are no explicit initial conditions. Instead, the initial conditions are included as additional terms $f \otimes \delta'_{t_0}$, $g \otimes \delta_{t_0}$ in the source term k.

The Existence and Uniqueness of Distribution Solutions. The goal of this article is to present an elementary and constructive proof of

THEOREM 4. *The Cauchy problem* (14), (15) *has a unique distribution solution* $u \in \mathscr{D}'(R^3)$ *for every distribution* $k \in \mathscr{D}'(R^3)$ *that satisfies* (17).

Proof of Uniqueness. The strategy of the proof is to show that any solution $u \in \mathscr{D}'(R^3)$ of (14), (15) must be given by a suitable version of Parseval's formula (3). With this in mind it is helpful to consider the functions

$$\begin{aligned} u_\phi(x, y, t) &= u(\phi(x' - x, y' - y, t' - t)) \\ k_\phi(x, y, t) &= k(\phi(x' - x, y' - y, t' - t)), \end{aligned} \tag{18}$$

where $\phi \in \mathscr{D}(R^3)$ and u and k act on the variables x', y', t'. u_ϕ and k_ϕ are the convolutions of u and k, respectively, with the testing function $\phi(-x, -y, -t)$. Hence a standard result of distribution theory implies that $u_\phi, k_\phi \in C^\infty(R^3)$ and differentiations with respect to the variables x, y, t commute with the actions of u and k; see, for example, [5, p: 14]. These properties are also easy to verify directly.

u_ϕ and k_ϕ may be thought of as smoothed versions of u and k. It will be shown that u_ϕ is a classical solution of the Cauchy problem

$$(\partial^2/\partial t^2 - \partial^2/\partial x^2 - \partial^2/\partial y^2)u_\phi = k_\phi \qquad \forall\, (x, y, t) \in R^2 \times [t_1, \infty), \tag{19}$$

$$u_\phi(x, y, t_1) = 0, \qquad (\partial/\partial t)u_\phi(x, y, t_1) = 0 \qquad \forall\, (x, y) \in R^2, \tag{20}$$

where $t_1 = t_1(\phi) < 0$ is a suitable number. To verify these equations note that the rules of distributional calculus and equation (14) imply that

$$\begin{aligned} &(\partial^2/\partial t^2 - \partial^2/\partial x^2 - \partial^2/\partial y^2)u_\phi(x, y, t) \\ &\quad = u\big((\partial^2/\partial t^2 - \partial^2/\partial x^2 - \partial^2/\partial y^2)\phi(x' - x, y' - y, t' - t)\big) \\ &\quad = u\big((\partial^2/\partial t'^2 - \partial^2/\partial x'^2 - \partial^2/\partial y'^2)\phi(x' - x, y' - y, t' - t)\big) \\ &\quad = (u_{tt} - u_{xx} - u_{yy})(\phi(x' - x, y' - y, t' - t)) \\ &\quad = k(\phi(x' - x, y' - y, t' - t)), \qquad \forall\, (x, y, t) \in R^3, \end{aligned} \tag{21}$$

which implies (19). Next, note that

$$(x' - x, y' - y, t' - t) \in \operatorname{supp} \phi \Leftrightarrow (x', y', t') \in (x, y, t) + \operatorname{supp} \phi. \tag{22}$$

Thus if $T(\phi) = \operatorname{Max}\{t' | (x', y', t') \in \operatorname{supp} \phi\}$ then

$$T(\phi) + t = \operatorname{Max}\{t' | (x', y', t') \in \operatorname{supp} \phi(x' - x, y' - y, t' - t)\}. \tag{23}$$

In particular, (15) implies that $\operatorname{supp} u \cap \operatorname{supp} \phi(x' - x, y' - y, t' - t) = \varnothing$ when $T(\phi) + t < t_0$ or $t < t_0 - T(\phi)$. Hence, if $t_1 = t_0 - T(\phi)$ then $\operatorname{supp} u_\phi \subset R^2 \times [t_1, \infty)$ and similarly $\operatorname{supp} k_\phi \subset R^2 \times [t_1, \infty)$. In particular, (20) holds. Note that it may be assumed that $T(\phi) > t_0$, and hence $t_1 < 0$, since otherwise u_ϕ and k_ϕ are identically zero.

Equations (19), (20) and Theorems 1 and 2 imply that u_ϕ and k_ϕ are necessarily related by Parseval's formula:

$$u_\phi(x, y, t) = \frac{1}{2\pi} \int_{t_1}^{t} \iint_{D(x, y, t-\tau)} \frac{k_\phi(\xi, \eta, \tau)}{\sqrt{(t - \tau)^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta\, d\tau \tag{24}$$

for all $(x, y, t) \in R^2 \times [t_1, \infty)$. In particular, since $t_1 < 0$ and $u_\phi(0, 0, 0) = u(\phi)$, by

(18), one has

$$u(\phi) = \frac{1}{2\pi}\int_{t_1}^{0}\iint_{D(0,0,-\tau)} \frac{k_\phi(\xi,\eta,\tau)}{\sqrt{\tau^2-\xi^2-\eta^2}}\,d\xi\,d\eta\,d\tau. \tag{25}$$

Thus any solution u of (14), (15) is related to the source term k by (25). This completes the proof of the uniqueness statement of Theorem 4.

Proof of Existence. It is natural to use (25) to construct a distribution solution of the Cauchy problem (14), (15). Recall that in (25) $t_1 = t_0 - T(\phi)$ depends on ϕ. However, since $\operatorname{supp} k_\phi \subset R^2 \times [t_0, \infty)$, one can write

$$u(\phi) = \frac{1}{2\pi}\int_{-\infty}^{0}\iint_{D(0,0,-\tau)} \frac{k_\phi(\xi,\eta,\tau)}{\sqrt{\tau^2-\xi^2-\eta^2}}\,d\xi\,d\eta\,d\tau. \tag{26}$$

The existence proof will be completed by showing that, for any $k \in \mathscr{D}'(R^3)$ with $\operatorname{supp} k \subset R^2 \times [t_0, \infty)$, equation (26) defines a distribution $u \in \mathscr{D}'(R^3)$ that satisfies (14) and (15).

Proof that (26) defines a distribution. For each $\phi \in \mathscr{D}(R^3)$ one has $k_\phi \in C^\infty(R^3)$ and $\operatorname{supp} k_\phi \subset R^2 \times [t_1, \infty)$. Hence, $\forall\ \phi \in \mathscr{D}(R^3)$, (26) defines a finite number $u(\phi)$. Moreover, the linearity of $u(\phi)$ is evident from (18) and (26). It remains to show that $u(\phi)$ is continuous in the distribution sense. This means that to each compact set $K \subset R^3$ there must correspond constants $C = C(K) \geq 0$ and $N = N(K) \geq 0$ such that one has [5, p. 4]

$$|u(\phi)| \leq C \sum_{\alpha_1+\alpha_2+\alpha_3 \leq N} \sup_K |D_x^{\alpha_1} D_y^{\alpha_2} D_t^{\alpha_3}\phi| \tag{27}$$

for all $\phi \in \mathscr{D}(R^3)$ with $\operatorname{supp}\phi \subset K$. To verify this define $T(K) = \operatorname{Max}\{t | (x, y, t) \in K\}$ and assume that $T(K) > t_0$ (since otherwise $u(\phi) \equiv 0$). Then $t_1(K) = t_0 - T(K) < 0$ and $\forall\ \phi \in \mathscr{D}(R^3)$ with $\operatorname{supp}\phi \subset K$ one has $\operatorname{supp} k_\phi \subset R^2 \times [t_1(K), \infty)$ whence

$$u(\phi) = \frac{1}{2\pi}\iiint_{K'} \frac{k_\phi(\xi,\eta,\tau)}{\sqrt{\tau^2-\xi^2-\eta^2}}\,d\xi\,d\eta\,d\tau, \tag{28}$$

where

$$K' = \left\{(\xi,\eta,\tau) | \tau + \sqrt{\xi^2+\eta^2} \leq 0 \text{ and } t_1(K) \leq \tau \leq 0\right\} \tag{29}$$

is the solid cone with vertex at the origin and base $D(0, 0, -t_1(K))$ in the plane $\tau = t_1(K) < 0$. Equation (28) will be used to prove (27).

The sets K and K' are compact and so is their vector sum

$$K + K' = \{(x + x', y + y', t + t') | (x, y, t) \in K \text{ and } (x', y', t') \in K'\}. \tag{30}$$

Hence, since $k \in \mathscr{D}'(R^3)$ by hypothesis, $\exists$ constants $C' = C(K + K')$, $N' = N(K + K')$ such that

$$|k(\phi)| \leq C' \sum_{\alpha_1+\alpha_2+\alpha_3 \leq N'} \sup_{K+K'} |D_x^{\alpha_1} D_y^{\alpha_2} D_t^{\alpha_3}\phi| \tag{31}$$

for all $\phi \in \mathscr{D}(R^3)$ with $\operatorname{supp}\phi \subset K + K'$. In particular, if $\operatorname{supp}\phi \subset K$ and $(\xi, \eta, \tau) \in K'$ then $(x' - \xi, y' - \eta, t' - \tau) \in K \Rightarrow (x', y', t') \in K + K'$. Thus

$\phi(x' - \xi, y' - \eta, t' - \tau)$, as a function of (x', y', t') with $(\xi, \eta, \tau) \in K'$ fixed, has support in $K + K'$ and (31) is valid with ϕ replaced by this function. Hence (31) and (18) imply that

$$\begin{aligned} \left|k_\phi(\xi,\eta,\tau)\right| &= \left|k(\phi(x' - \xi, y' - \eta, t' - \tau))\right| \\ &\le C' \sum_{\alpha_1+\alpha_2+\alpha_3 \le N'} \sup_{K+K'} \left|D_x^{\alpha_1} D_y^{\alpha_2} D_t^{\alpha_3} \phi(x - \xi, y - \eta, t - \tau)\right|, \end{aligned} \tag{32}$$

provided that $\operatorname{supp} \phi \subset K$ and $(\xi, \eta, \tau) \in K'$. Moreover, for each fixed $(\xi, \eta, \tau) \in K'$, the support of $\phi(x - \xi, y - \eta, t - \tau)$, as a function of (x, y, t), lies in $(\xi, \eta, \tau) + K \subset K' + K$, whence (32) can be written

$$\begin{aligned} \left|k_\phi(\xi,\eta,\tau)\right| &\le C' \sum_{\alpha_1+\alpha_2+\alpha_3 \le N'} \sup_{(\xi,\eta,\tau)+K} \left|D_x^{\alpha_1} D_y^{\alpha_2} D_t^{\alpha_3} \phi(x - \xi, y - \eta, t - \tau)\right| \\ &= C' \sum_{\alpha_1+\alpha_2+\alpha_3 \le N'} \sup_{K} \left|D_x^{\alpha_1} D_y^{\alpha_2} D_t^{\alpha_3} \phi(x, y, t)\right| \end{aligned} \tag{33}$$

$\forall\, \phi \in \mathscr{D}(R^3)$ with $\operatorname{supp} \phi \subset K$ and $\forall\, (\xi, \eta, \tau) \in K'$. Relations (28) and (33) imply (27). To see this note that the right hand side of (28) with $k_\phi(\xi, \eta, \tau) \equiv 1$ is an elementary integral with value $(1/2)t_1^2(K)$. Thus, (28) and (31) imply

$$|u(\phi)| \le \frac{1}{2} t_1^2 C' \sum_{\alpha_1+\alpha_2+\alpha_3 \le N'} \sup_K |D_x^{\alpha_1} D_{yt}^{\alpha_2 \alpha_3} \phi|, \tag{34}$$

which is equivalent to (27).

Proof that u satisfies (14). It must be shown that $\forall\, \phi \in \mathscr{D}(R^3)$ one has $(u_{tt} - u_{xx} - u_{yy})(\phi) = k(\phi)$. Now the definition (28) and the rules of distributional calculus give

$$\begin{aligned} (u_{tt} - u_{xx} - u_{yy})(\phi) &= u(\phi_{tt} - \phi_{xx} - \phi_{yy}) \\ &= \frac{1}{2\pi} \iiint_{K'} \frac{l_\phi(\xi,\eta,\tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}}\, d\xi\, d\eta\, d\tau, \end{aligned} \tag{35}$$

where l_ϕ is given by

$$\begin{aligned} l_\phi(\xi,\eta,\tau) &= k_{\phi_{tt} - \phi_{xx} - \phi_{yy}}(\xi,\eta,\tau) \\ &= k\big((\partial^2/\partial t^2 - \partial^2/\partial x^2 - \partial^2/\partial y^2)\phi(x - \xi, y - \eta, t - \tau)\big) \\ &= k\big((\partial^2/\partial \tau^2 - \partial^2/\partial \xi^2 - \partial^2/\partial \eta^2)\phi(x - \xi, y - \eta, t - \tau)\big) \\ &= (\partial^2/\partial \tau^2 - \partial^2/\partial \xi^2 - \partial^2/\partial \eta^2) k_\phi(\xi,\eta,\tau). \end{aligned} \tag{36}$$

In particular, $l_\phi \in C^\infty(R^3)$ and $\operatorname{supp} l_\phi \subset R^2 \times [t_1, \infty)$. Thus, k_ϕ is the classical solution of the Cauchy problem

$$(\partial^2/\partial \tau^2 - \partial^2/\partial \xi^2 - \partial^2/\partial \eta^2) k_\phi = l_\phi \quad \forall\, (\xi,\eta,\tau) \in R^2 \times [t_1, \infty), \tag{37}$$

$$k_\phi(\xi,\eta,t_1) = 0, \qquad (\partial/\partial t) k_\phi(\xi,\eta,t_1) = 0 \qquad \forall\, (\xi,\eta) \in R^2.$$

and Theorems 1 and 2 imply that

$$k_\phi(x,y,t) = \frac{1}{2\pi}\int_{t_1}^{t}\iint_{D(x,y,t-\tau)} \frac{l_\phi(\xi,\eta,\tau)}{\sqrt{(t-\tau)^2-(x-\xi)^2-(y-\eta)^2}}\,d\xi\,d\eta\,d\tau \tag{38}$$

for all $(x, y, t) \in R^2 \times [t_1, \infty)$. Finally, combining (35) and (38) gives

$$(u_{tt} - u_{xx} - u_{yy})(\phi) = k_\phi(0,0,0) = k(\phi) \qquad \forall\, \phi \in \mathscr{D}(R^3). \tag{39}$$

Thus u satisfies (14).

Proof that u satisfies (15). It must be shown that if $\phi \in \mathscr{D}(R^3)$ then $u(\phi) = 0$ whenever supp $\phi \subset R^2 \times (-\infty, t_0)$. But in this case one has $T(\phi) = \text{Max}\{t | (x, y, t) \in \text{supp } \phi\} \leq t_0$ and hence one may take $t_1 = t_0 - T(\phi) = 0$. This gives $u(\phi) = 0$ by (25).

This completes the proof of Theorem 4.

The Fundamental Solution. A fundamental solution for the wave equation is a distribution solution of the wave equation

$$u_{tt} - u_{xx} - u_{yy} = \delta, \tag{40}$$

where δ is the Dirac delta distribution in R^3, defined by

$$\delta(\phi) = \phi(0,0,0) \qquad \forall\, \phi \in \mathscr{D}(R^3). \tag{41}$$

Theorem 4 implies that there is a unique fundamental solution G that has its support in the half-space $R^2 \times [0, \infty)$. Thus

$$G_{tt} - G_{xx} - G_{yy} = \delta \in \mathscr{D}'(R^3), \tag{42}$$

$$\text{supp } G \subset R^2 \times [0, \infty). \tag{43}$$

In some texts G is called the *Green's function*, or the *Riemann function*, of the wave equation. Here G will be constructed explicitly and used to represent the solution of the distributional Cauchy problem (14), (15).

The *forward* and *backward wave cones* with vertices at the origin will be denoted by Γ_+ and Γ_-, respectively. Thus

$$\begin{aligned} \Gamma_+ &= \left\{(x,y,t) | t - \sqrt{x^2+y^2} \geq 0\right\}, \\ \Gamma_- &= \left\{(x,y,t) | t + \sqrt{x^2+y^2} \leq 0\right\}. \end{aligned} \tag{44}$$

Note that equation (26) which defines the solution of the Cauchy problem can be written

$$u(\phi) = \frac{1}{2\pi}\iiint_{\Gamma_-} \frac{k(\phi(x-\xi, y-\eta, t-\tau))}{\sqrt{\tau^2-\xi^2-\eta^2}}\,d\xi\,d\eta\,d\tau, \tag{45}$$

where k acts on the variables (x, y, t). Of course the integrand in (45) vanishes

outside a compact subset of Γ_-. On taking $k = \delta$ one gets

$$G(\phi) = \frac{1}{2\pi}\iiint_{\Gamma_-} \frac{\phi(-\xi, -\eta, -\tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}}\, d\xi\, d\eta\, d\tau$$
$$= \frac{1}{2\pi}\iiint_{\Gamma_+} \frac{\phi(\xi, \eta, \tau)}{\sqrt{\tau^2 - \xi^2 - \eta^2}}\, d\xi\, d\eta\, d\tau, \tag{46}$$

since Γ_- is carried into Γ_+ by the mapping $(\xi, \eta, \tau) \to (-\xi, -\eta, -\tau)$. This result shows that the distribution G is actually a locally integrable function with $\operatorname{supp} G = \Gamma_+$. It may be written

$$G(x, y, t) = \frac{1}{2\pi} \frac{H\left(t - \sqrt{x^2 + y^2}\right)}{\sqrt{t^2 - x^2 - y^2}}, \tag{47}$$

where $H(t - \sqrt{x^2 + y^2})$ is the characteristic function of Γ_+.

The principle of superposition for the Cauchy problem is expressed by

THEOREM 5. *$\forall\ k \in \mathscr{D}'(R^3)$ with $\operatorname{supp} k \subset R^2 \times [t_0, \infty)$ the solution u of the Cauchy problem* (14), (15) *satisfies*

$$u = G * k. \tag{48}$$

Proofs of equation (48) are given in most standard texts; see [5], p. 139 or [1], Ch. 5. It may also be verified directly from equation (45) by using the easily verified relation

$$(G * k)(\phi) = (G * k_\phi)(0, 0, 0). \tag{49}$$

Concluding Remarks. A great advantage of the distributional formulation of the Cauchy problem is that it permits a definitive discussion of the existence, uniqueness and construction of the solution without *a priori* assumptions about the Cauchy data f, g, h. Questions concerning the regularity of the solution are then secondary, to be answered by means of the solution formulas (45) or (48). As a concluding example, conditions will be given which guarantee that the original Parseval formula (3) define a locally integrable solution. Parseval's formula with $f = g = 0$ can be written

$$u(x, y, t) = \frac{1}{2\pi}\iiint_{\Gamma_-(x, y, t)} \frac{h(\xi, \eta, \tau)}{\sqrt{(t - \tau)^2 - (x - \xi)^2 - (y - \eta)^2}}\, d\xi\, d\eta\, d\tau, \tag{50}$$

where $\Gamma_-(x, y, t)$ is the backward wave cone with vertex (x, y, t), defined by

$$\Gamma_-(x, y, t) = \left\{(\xi, \eta, \tau) \mid (\tau - t) + \sqrt{(x - \xi)^2 + (y - \eta)^2} \le 0\right\}.$$

Formally, (50) is just the convolution equation (48). Now it is easy to verify by direct integration that $G(x, y, t) \in L_p^{\text{loc}}(R^3)$ for $1 \le p < 2$, where

$$L_p^{\text{loc}} = \left\{f(x, y, t) \mid \iiint_K |f(x, y, t)|^p\, dx\, dy\, dt < \infty \qquad \forall \text{ compact } K \subset R^3\right\}.$$

It follows by Hölder's inequality that if $h \in L_q^{\text{loc}}(R^3)$ with $2 < q \leq +\infty$ (and supp $h \subset R^2 \times [t_0, \infty)$) then the integral in (50) converges absolutely for all $(x, y, t) \in R^3$. In this case the function $u(x, y, t)$ defined by (50) is in $L_1^{\text{loc}}(R^3)$ and coincides with the distribution solution of the Cauchy problem. A proof may be obtained from equation (45) by writing

$$k(\phi(x - \xi, y - \eta, t - \tau)) = \iiint_{R^3} h(x, y, t)\phi(x - \xi, y - \eta, t - \tau)\, dx\, dy\, dt$$

and using Fubini's theorem to write $u(\phi)$ as

$$u(\phi) = \iiint_{R^3} u(x, y, t)\phi(x, y, t)\, dx\, dy\, dt.$$

A similar argument shows that Parseval's formula with $g \in L_q^{\text{loc}}(R^2)$, $2 < q \leq +\infty$, and $f = 0$, $h = 0$ defines a function $u_{(0,g,0)}(x, y, t) \in L_1^{\text{loc}}(R^3)$ which coincides with the distribution solution with source term $k = g \otimes \delta_{t_0}$. If $f \in L_q^{\text{loc}}(R^2)$, $2 < q \leq +\infty$, and $g = 0$, $h = 0$, one can only say that $u_{(f,0,0)}$ is the distributional t-derivative of a function $P_{x,y,t-t_0}(f)$ in $L_1^{\text{loc}}(R^3)$.

REFERENCES

1. F. G. Friedlander, Introduction to the Theory of Distributions, Cambridge University Press, Cambridge, 1982.
2. P. Garabedian, Partial Differential Equations, J. Wiley and Sons, New York, 1964.
3. I. M. Gelfand and G. E. Shilov, Generalized Functions, vol. 1, Academic Press, New York, 1964.
4. J. Hadamard, Lectures on Cauchy's Problem, Dover Publ. Co., New York, 1952.
5. L. Hörmander, Linear Partial Differential Operators, Springer Verlag, New York, 1963.
6. F. John, Partial Differential Equations, fourth edition, Springer Verlag, New York, 1982.
7. L. Schwartz, Théorie des Distributions, vol. 1 and 2, Hermann et Cie., Paris, 1950–51.
8. A. G. Webster, Partial Differential Equations of Mathematical Physics, Dover Publ. Co., New York, 1955.
9. E. C. Zachmanoglou and D. W. Thoe, Introduction to Partial Differential Equations with Applications, Dover Publ. Co., New York, 1988.

Continuous Nowhere-Differentiable Functions— an Application of Contraction Mappings

HIDEFUMI KATSUURA, *San Jose State University*

HIDEFUMI KATSUURA was born and raised in Wakayama about 300 miles south of Tokyo, Japan. He came to the United States to study at the College of Charleston, where he received his B.A. He earned his Ph.D. at the University of Delaware under Professor David P. Bellamy. Currently, he is teaching and doing research in point set topology and in analysis.



1. Introduction. The existence of a continuous nowhere-differentiable function is well known, and many such functions are known. Most of them are variations of the following examples.

Examples of nowhere differentiable functions:

$$\text{If } f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n}\cos(3^n x), \tag{1}$$

then f is a continuous nowhere-differentiable function. (This is due to Weierstrass. See [2, p. 195].)

(2) Let $g(x)$ be the distance from x to the nearest integer. If

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n} g(2^n x),$$

then f is also a continuous nowhere-differentiable function. (See [3, p. 115].)

The above examples have concise definitions and establish the existence of continuous nowhere-differentiable functions. However, it is not easy to visualize or guess what their graphs look like, let alone to see intuitively why they work. Our continuous nowhere-differentiable function $\mathbf{f}$: $[0, 1] \to \mathbf{R}$ is the uniform limit of a sequence of piecewise linear continuous functions f_n: $[0, 1] \to \mathbf{R}$ with steep slopes. In constructing the sequence $\langle f_n \rangle$, we will be using a contraction mapping w from the family of all closed subsets of $X = [0, 1] \times [0, 1]$ into itself with respect to the Hausdorff metric induced by the Euclidean metric. (A contraction mapping on a metric space (Y, d) is a function g: $Y \to Y$ for which there is a positive constant $k < 1$ such that $d(g(x), g(y)) \leq kd(x, y)$ for every $x, y \in Y$.) It is interesting to note that, in our example, if A is *any non-empty closed subset* of the square X, even if it is a singleton set, the iterated sequence $\langle w^n(A) \rangle$ of sets converges to the graph of $\mathbf{f}$ in the Hausdorff metric. If A is the diagonal of slope 1 in the square X, then $w^n(A)$ is the graph of f_n, and hence, a reader can obtain an intuitive idea of the graph of $\mathbf{f}$. This idea first occurred to me when I attended Mr. Gary Church's master's thesis defense at San Jose State University at which he talked about attractors of contraction mappings (see [4]).

II. The construction of a continuous nowhere-differentiable function. Let $X = [0, 1] \times [0, 1]$ be the closed unit square in the plane with the usual Euclidean

metric. Let w_i: $X \to X$, $i = 1, 2, 3$ be contraction mappings defined by

$$w_1(x, y) = \left(\frac{x}{3}, \frac{2y}{3}\right),$$

$$w_2(x, y) = \left(\frac{2 - x}{3}, \frac{1 + y}{3}\right),$$

and

$$w_3(x, y) = \left(\frac{2 + x}{3}, \frac{1 + 2y}{3}\right).$$

(See FIGURE 1.)

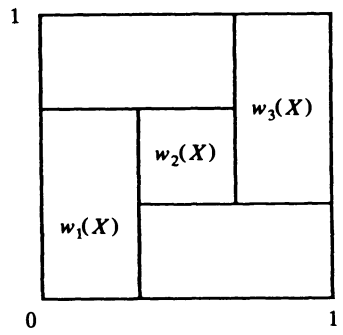


FIG. 1

In FIGURE 1, w_1 shrinks X onto $w_1(X)$, with the fixed point $(0, 0)$, w_2 flips X about the line $x = 1/2$ and shrinks it onto $w_2(X)$, with the fixed point $(1/2, 1/2)$, and w_3 shrinks X onto $w_3(X)$, with the fixed point $(1, 1)$.

Let $\mathbf{F}(X)$ be the collection of all non-empty closed subsets of X. We define a function w from $\mathbf{F}(X)$ into itself by

$$w(A) = w_1(A) \cup w_2(A) \cup w_3(A)$$

for every A in $\mathbf{F}(X)$. For every A, B in $\mathbf{F}(X)$, let

$$d_H(A, B) = \inf\{\varepsilon > 0: N_\varepsilon(A) \supset B \text{ and } N_\varepsilon(B) \supset A\},$$

where $N_\varepsilon(A)$ and $N_\varepsilon(B)$ are ε-neighborhoods of A and B, respectively. Then d_H is a complete metric on $\mathbf{F}(X)$, and is called a *Hausdorff metric*. It is interesting to note that the function w is a contraction mapping on $\mathbf{F}(X)$ under this metric d_H (see [5]).

Let $D_0 = \{(x, x) \in X\}$ be the diagonal in X. For each $n = 1, 2, 3, \ldots$, let $D_n = w(D_{n-1})$. Then for every $n = 0, 1, 2, \ldots$, D_n is the graph of a continuous function, call it f_n, from $[0, 1]$ onto itself.

Note that if $m \leq n$, then $w^n(X) \supseteq D_m$, and $w^n(X)$ is the union of 3^n rectangles of heights $\leq (2/3)^n$. Hence, we have

$$\sup\{|f_m(t) - f_n(t)|: t \in [0, 1]\} \leq (2/3)^n.$$

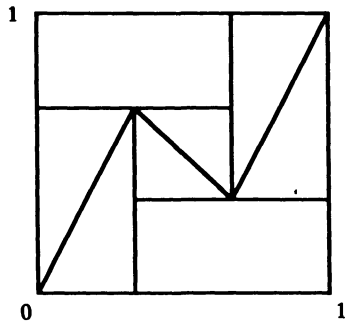


FIG. 2. $w(X)$ and D_1.

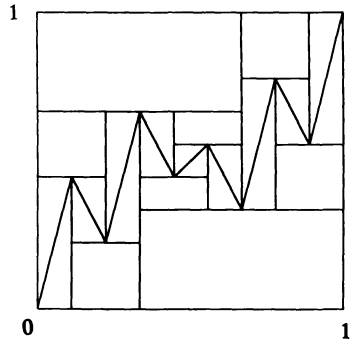


FIG. 3. $w^2(X)$ and D_2.

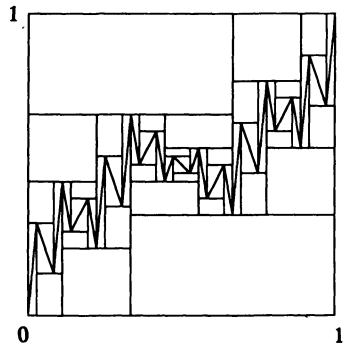


FIG. 4. $w^3(X)$ and D_3.

This implies that the sequence of continuous functions $\langle f_n \rangle$ is uniformly Cauchy. Thus there is a continuous function **f** from [0, 1] onto itself such that f_n converges to **f** uniformly on [0, 1]. We will prove that this function **f** is nowhere-differentiable on the interval (0, 1).

Since w is a contraction mapping on $\mathbf{F}(X)$, it has a unique fixed point D in $\mathbf{F}(X)$ (see Theorem 4.48 on page 92, [1]). Note that if A is an arbitrary element of $\mathbf{F}(X)$, then the sequence $\langle w^n(A) \rangle$ converges to D with respect to the metric d_H. Hence, D_n converges to D and D is the graph of **f**. If we let $B = \{(0, 0), (1/2, 1/2), (1, 1)\}$ be the set of fixed points of w_1, w_2, and w_3, then the sequence $\langle w^n(B) \rangle$ in $\mathbf{F}(X)$ also converges to D. Moreover, $D \supseteq w^n(B)$. Hence, if one plots the points of $w^n(B)$ for a large n, then not only does it resemble the graph of **f**, but actually it is a part of the graph of **f**.

III. Proof of the nowhere-differentiability of f. Let T be the set of all ternary rationals in the interval (0, 1), and for each $n = 1, 2, 3, \ldots$, let T_n be the set of n-digit ternary rationals.

LEMMA 1. *Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be sequences in T such that, for each $n = 1, 2, \ldots$,*
(1) *x_n and y_n are in T_n,*
(2) *$y_n - x_n = 1/3^n$, and*
(3) *either $x_n = x_{n+1}$ or $y_n = y_{n+1}$.*

Then

$$\lim_{n\to\infty}\left|\frac{\mathbf{f}(y_n)-\mathbf{f}(x_n)}{y_n-x_n}\right|=\infty.$$

Proof. We claim that

$$\left|\frac{\mathbf{f}(y_n)-\mathbf{f}(x_n)}{y_n-x_n}\right|\geq 2^{n-1}\qquad (n=1,2,3,\dots).$$

First, note that if $x\in T_n$, then $\mathbf{f}(x)=f_n(x)$. If $n=1$, then

$$\frac{\mathbf{f}(y_1)-\mathbf{f}(x_1)}{y_1-x_1}=\begin{cases}\dfrac{\frac{2}{3}}{\frac{1}{3}}=2 & \text{if } \left(x_1=0, \text{ and } y_1=\tfrac{1}{3}\right) \text{ or } \left(x_1=\tfrac{2}{3} \text{ and } y_1=1\right),\\[2ex] \dfrac{-\frac{1}{3}}{\frac{1}{3}}=-1 & \text{if } x_1=\tfrac{1}{3} \text{ and } y_1=\tfrac{2}{3}.\end{cases}$$

Hence, if $n=1$, we have proved the claim.

Suppose, for some integer $k\geq 1$, the claim is true, i.e.,

$$|\mathbf{f}(y_k)-\mathbf{f}(x_k)|\geq 2^{k-1}\cdot|y_k-x_k|.$$

Then

$$\left|\frac{\mathbf{f}(y_{k+1})-\mathbf{f}(x_{k+1})}{y_{k+1}-x_{k+1}}\right|=\frac{\frac{2}{3}|\mathbf{f}(y_k)-\mathbf{f}(x_k)|}{\frac{1}{3}|y_k-x_k|}$$

(see FIGURE 5 for the case $x_k=x_{k+1}$)

$$\geq 2\cdot\frac{2^{k-1}|y_k-x_k|}{|y_k-x_k|}$$

by the induction hypothesis

$$=2^{(k+1)-1}$$

which proves the lemma, by induction. ◆

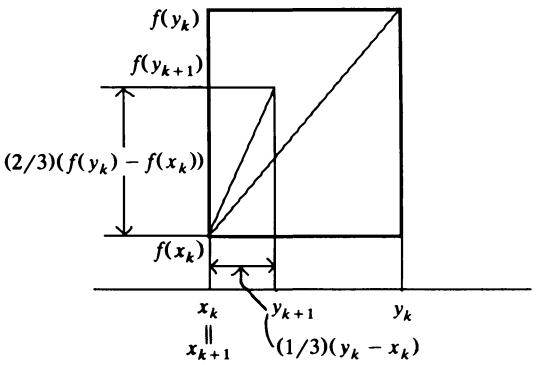


FIG. 5

LEMMA 2. *Let* $\langle x_n \rangle$ *and* $\langle y_n \rangle$ *be sequences in* T *such that for each* $n = 1, 2, \ldots, x_n, y_n \in T_n$, *and* $y_n - x_n = 1/3^n$. *Suppose* $x_n \neq x_{n+1}$ *and* $y_n \neq y_{n+1}$ *for infinitely many* $n = 1, 2, \ldots$. *Then*

$$\lim_{n \to \infty} \frac{\mathbf{f}(y_n) - \mathbf{f}(x_n)}{y_n - x_n}$$

does not exist.

Proof. First, we claim that

$$\left| \frac{\mathbf{f}(y_n) - \mathbf{f}(x_n)}{y_n - x_n} \right| \geq 1 \quad \text{for each } n = 1, 2, 3, \ldots$$

From the proof of the previous lemma, this is surely true if $n = 1$. First, if $x_k = x_{k+1}$ or $y_k = y_{k+1}$, then as in the previous lemma, we have

$$\left| \frac{\mathbf{f}(y_{k+1}) - \mathbf{f}(x_{k+1})}{y_{k+1} - x_{k+1}} \right| = \frac{\frac{2}{3}|\mathbf{f}(y_k) - \mathbf{f}(x_k)|}{\frac{1}{3}|y_k - x_k|}$$

(see FIGURE 5 for the case $x_k = x_{k+1}$)

$$\geq 2 \cdot 1$$

by the induction hypothesis

$$> 1.$$

On the other hand, if $x_k \neq x_{k+1}$ and $y_k \neq y_{k+1}$, then

$$\left| \frac{\mathbf{f}(y_{k+1}) - \mathbf{f}(x_{k+1})}{y_{k+1} - x_{k+1}} \right| = \frac{\frac{1}{3}|\mathbf{f}(y_k) - \mathbf{f}(x_k)|}{\frac{1}{3}|y_k - x_k|}$$

(see FIGURE 6)

$$\geq 1$$

by the induction hypothesis, which proves the claim.

Now, if n is an integer such that $x_n \neq x_{n+1}$ and $y_n \neq y_{n+1}$, then

$$\frac{\mathbf{f}(y_{n+1}) - \mathbf{f}(x_{n+1})}{y_{n+1} - x_{n+1}} = \frac{-\frac{1}{3}(\mathbf{f}(y_n) - \mathbf{f}(x_n))}{\frac{1}{3}(y_n - x_n)}$$

(see FIGURE 6)

$$= -\left(\frac{\mathbf{f}(y_n) - \mathbf{f}(x_n)}{y_n - x_n} \right).$$

So surely if the limit exists it must be 0. But by the claim proved above, it can't be 0 either. This means that the limit does not exist. ♦

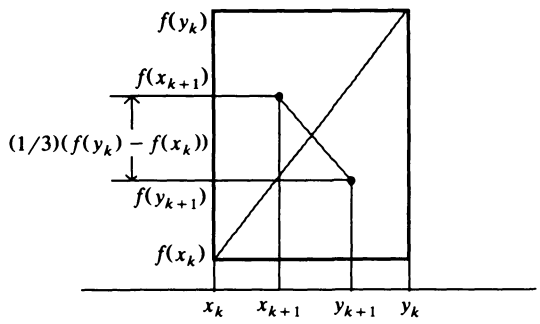


FIG. 6

The following is well known.

LEMMA 3. *Let g be a real valued function on* $[0, 1]$ *differentiable at* t, $0 < t < 1$. *If, for every* $n = 1, 2, 3, \ldots$, $0 < x_n < t < y_n < 1$, *and if* $x_n \to t$ *and* $y_n \to t$ *then,*

$$\lim_{n \to \infty} \frac{g(y_n) - g(x_n)}{y_n - x_n} = g'(t).$$

THEOREM. *The function* $\mathbf{f}$: $[0, 1] \to [0, 1]$ *is not differentiable at any* $x \in (0, 1)$.

Proof. If $x \in T_n$ for some integer n, then let $y_k = x + 1/3^{n+k}$ for every $k = 1, 2, \ldots$. Then $\lim_{k \to \infty} y_k = x$ and

$$\lim_{k \to \infty} \left| \frac{\mathbf{f}(y_k) - \mathbf{f}(x_k)}{y_k - x_k} \right| = \infty$$

by Lemma 1. Hence, $\mathbf{f}$ is not differentiable at x.

On the other hand, if $x \notin T$, then there are sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ such that, for each $n = 1, 2, \ldots$,

(1) x_n and y_n are both in T_n,

(2) $y_n - x_n = 1/3^n$, and

(3) $a_n < x < b_n$.

Then by Lemmas 1 and 2, we know that

$$\lim_{n \to \infty} \frac{\mathbf{f}(y_n) - \mathbf{f}(x_n)}{y_n - x_n}$$

does not exist. Hence, by Lemma 3, $\mathbf{f}$ is not differentiable at x. ♦

REFERENCES

1. Tom M. Apostol, Mathematical Analysis, Addison-Wesley Publishing Company, 1974.
2. Robert G. Bartle, The Elements of Real Analysis, John Wiley and Sons, Inc., 1976.
3. Ralph P. Boas, Jr., A Primer of Real Functions, The Math. Asso. of Amer., 1960.
4. Gary Church, Computer Approximation of Fractal Sets using Iterated Function Systems, Master's Thesis, San Jose State University, San Jose, California, 1988.
5. J. Hutchinson, Fractals and Self Similarity, *Indiana Univ. Math. J.*, 30 (1981) 713–747.

LETTERS TO THE EDITOR

Correction to "Overdetermined Systems of Linear Equations" (This MONTHLY, 97 (1990) 511–513)

The result announced in this paper is false. It says that "the least squares solution to an overdetermined system of linear equations $AX = Y$ is the point such that the sum of the squares of the distances from the point to each of the subsets defined by the equations is a minimum." This conclusion followed from the erroneous step $(A'D^2A)^{-1}A'D^2Y = (D^2A'A)^{-1}D^2A'Y$, where D was a positive diagonal matrix.

If the assertion were true, it would imply that the systems $AX = Y$ and $DAX = DY$, where D is any diagonal matrix, have the same least squares solution, but that is not true. However, there are interesting connections between two such systems.

The system $DAX = DY$ is commonly called a weighted least squares system, and the weights d_{ii} modify the least squares solution. Such row scaling operations defined by D are nevertheless useful. Let R be the residual vector $R = Y - AX$. Assigning a relatively large weight to the ith equation will cause the corresponding component of the residual vector to be smaller. Thus, if some components of the data vector Y are known with more absolute accuracy than others, one can introduce relatively larger weights to these components [2].

This example in fact thus provides an interesting illustration for a linear algebra class of a "solution" (least squares) to a linear system that is not invariant under elementary row operations of the augmented matrix, and where this lack of invariance turns out to be very useful.

REFERENCES

1. Gareth Williams, Overdetermined systems of linear equations, this MONTHLY, 97 (1990) 511–513.
2. Charles L. Lawson and Richard J. Hanson, Solving Least Squares Problems, Prentice-Hall, 1974, pp. 183–185.

Gareth Williams

To the Editor:

1. The article, "Conway's Challenge Sequence" (January 1991), opens with the statement that J. H. Conway "introduced" the sequence $1, 1, 2, 2, 3, 4, \ldots$ in "a lecture at AT&T Bell Labs [in July, 1988]". The author later notes (p. 19) that a generalization of Conway's sequence is the subject of *Monthly* Problem E3274. I think it should be noted that the appearance of E3274 (June 1988) preceded Conway's lecture.

2. Newman and Parsons (January 1988, p. 44) give a short proof that every sequence of reals has a monotone subsequence; they then say (p. 45), "We have been unable to find [the proof] elsewhere [than in Newman's 1982 book, *A Problem Seminar*]". My colleague, George W. Booth, has found a similar proof in W. Maak's *An Introduction to Modern Calculus*, p. 29 (Holt, Rinehart & Winston, 1963).

Sincerely yours,
David M. Bloom
Brooklyn College of CUNY

LEMMA 4. *Let* (i, j, A_{ij}) *and* (k, l, A_{kl}) *be two distinct entries of* H *having minimum and maximum values, respectively. By querying* A_{il} *and doing a constant amount of extra computation, we can reduce the size of* H *by one, while preserving properties* $P1$ *and* $P2$.

Proof. By property P1, $i \neq k$ and $j \neq l$. We say a row or column is *remaining* if it has a representative in H. We divide the analysis into three cases depending on the value of A_{il}.

Case 1. $A_{il} \leqslant A_{ij} \leqslant A_{kl}$. Any strict saddlepoint in column l is no larger than A_{il}. However, $A_{il} \leqslant A_{ij}$, so by Lemma 2 column l cannot contain a strict SP, and we can eliminate this column entirely. By Lemma 3, the only possible strict SP in row k is A_{kl}, which we have already ruled out. Consequently, we can delete the entry (k, l, A_{kl}) from H.

Case 2. $A_{ij} \leqslant A_{kl} \leqslant A_{il}$. This case is symmetric to case 1; here we eliminate row i and column j and, consequently, delete the entry (i, j, A_{ij}) from H.

Case 3. $A_{ij} < A_{il} < A_{kl}$. By Lemma 3, the only possible strict saddlepoints in column j or row k are A_{ij} and A_{kl}. The first inequality rules out A_{ij}; the second rules out A_{kl}. Hence, the column j and the row k can be eliminated. The row i and the column l, however, cannot be eliminated yet. We, therefore, delete (i, j, A_{ij}) and (k, l, A_{kl}) from H but insert (i, l, A_{il}). This preserves the properties of H while decreasing its size by one.

This completes the proof. ■

Once we have eliminated all but one entry as a possible saddlepoint, then to test whether this last entry really is a saddlepoint, we make comparisons with all other entries in its row and column; this requires $2n - 2$ additional queries. Finally, we can store H as a *min-max heap* so that the operations Delete-Min, Delete-Max, Find-Min, Find-Max and Insert can be performed in worst-case time $O(\log n)$ [1]. This proves our main result.

THEOREM 5. *Given an* $m \times n$ *matrix* A, *where* $m \geqslant n$, *we can determine whether* A *has a strict saddlepoint, and report such an entry, by querying* $O(m)$ *matrix entries and doing* $O(m \log n)$ *additional computation.*

REFERENCES

1. M. D. Atkinson, J.-R. Sack, N. Santoro, and T. Strothotte, Min-max heaps and generalized priority queues, *Comm. ACM*, 29 (1986) 996–1000.
2. D. C. Llewellyn, C. Tovey, and M. Trick, Finding saddlepoints of two-person, zero sum games, *Amer. Math. Monthly*, 95 (1988) 912–918.

Circumscribed Circles

ROBERT OSSERMAN
Mathematics Department, Stanford University, Stanford, CA 94305

In a paper giving a new derivation of the four-vertex theorem [1], I stated without proof some elementary lemmas about circumscribed circles. The arguments needed are familiar to those who work in the field, but are not completely obvious. In rethinking those arguments, I noticed that the statements hold in far

LEMMA 4. *Let* (i, j, A_{ij}) *and* (k, l, A_{kl}) *be two distinct entries of* H *having minimum and maximum values, respectively. By querying* A_{il} *and doing a constant amount of extra computation, we can reduce the size of* H *by one, while preserving properties* $P1$ *and* $P2$.

Proof. By property P1, $i \neq k$ and $j \neq l$. We say a row or column is *remaining* if it has a representative in H. We divide the analysis into three cases depending on the value of A_{il}.

Case 1. $A_{il} \leqslant A_{ij} \leqslant A_{kl}$. Any strict saddlepoint in column l is no larger than A_{il}. However, $A_{il} \leqslant A_{ij}$, so by Lemma 2 column l cannot contain a strict SP, and we can eliminate this column entirely. By Lemma 3, the only possible strict SP in row k is A_{kl}, which we have already ruled out. Consequently, we can delete the entry (k, l, A_{kl}) from H.

Case 2. $A_{ij} \leqslant A_{kl} \leqslant A_{il}$. This case is symmetric to case 1; here we eliminate row i and column j and, consequently, delete the entry (i, j, A_{ij}) from H.

Case 3. $A_{ij} < A_{il} < A_{kl}$. By Lemma 3, the only possible strict saddlepoints in column j or row k are A_{ij} and A_{kl}. The first inequality rules out A_{ij}; the second rules out A_{kl}. Hence, the column j and the row k can be eliminated. The row i and the column l, however, cannot be eliminated yet. We, therefore, delete (i, j, A_{ij}) and (k, l, A_{kl}) from H but insert (i, l, A_{il}). This preserves the properties of H while decreasing its size by one.

This completes the proof. ■

Once we have eliminated all but one entry as a possible saddlepoint, then to test whether this last entry really is a saddlepoint, we make comparisons with all other entries in its row and column; this requires $2n - 2$ additional queries. Finally, we can store H as a *min-max heap* so that the operations Delete-Min, Delete-Max, Find-Min, Find-Max and Insert can be performed in worst-case time $O(\log n)$ [1]. This proves our main result.

THEOREM 5. *Given an* $m \times n$ *matrix* A, *where* $m \geqslant n$, *we can determine whether* A *has a strict saddlepoint, and report such an entry, by querying* $O(m)$ *matrix entries and doing* $O(m \log n)$ *additional computation.*

REFERENCES

1. M. D. Atkinson, J.-R. Sack, N. Santoro, and T. Strothotte, Min-max heaps and generalized priority queues, *Comm. ACM*, 29 (1986) 996–1000.
2. D. C. Llewellyn, C. Tovey, and M. Trick, Finding saddlepoints of two-person, zero sum games, *Amer. Math. Monthly*, 95 (1988) 912–918.

Circumscribed Circles

ROBERT OSSERMAN
Mathematics Department, Stanford University, Stanford, CA 94305

In a paper giving a new derivation of the four-vertex theorem [1], I stated without proof some elementary lemmas about circumscribed circles. The arguments needed are familiar to those who work in the field, but are not completely obvious. In rethinking those arguments, I noticed that the statements hold in far

greater generality than is generally realized. Since I do not know of a reference, I thought it worthwhile to provide proofs.

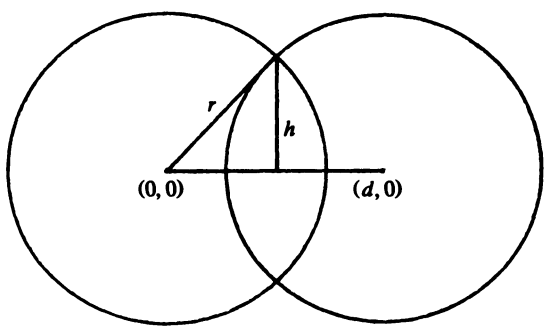
DEFINITION. For any set E in the plane, a *circumscribed circle* is a circle C such that E is included in the closed disk determined by C and not in any disk of smaller radius.

THEOREM 1. *Every bounded set E in the plane containing at least two points has a unique circumscribed circle.*

LEMMA 1. *Let C_k be a circle of radius r_k and center c_k, for $k = 1, 2$. Let $r = \max\{r_1, r_2\}$, and let d be the distance between c_1 and c_2. If $d < r_1 + r_2$, then the intersection of the disks D_1, D_2 corresponding to C_1, C_2 lies in a disk of radius h, where*

$$h^2 = r^2 - \left(\frac{d}{2}\right)^2. \tag{1}$$

Proof. Since $D_1 \cap D_2$ is included in the intersection of the disks of radius r centered at c_1, c_2, it is sufficient to prove the lemma when $r_1 = r_2 = r$. Then $d < 2r$, and the picture is



The assertion is clear from the picture and also follows immediately from the equations; assuming the centers are at $(0, 0)$ and $(d, 0)$, a point (x, y) in $D_1 \cap D_2$ must satisfy

$$x^2 + y^2 \leqslant r^2$$

and

$$(x - d)^2 + y^2 \leqslant r^2.$$

Adding the inequalities, we find that (x, y) satisfies

$$\left(x - \frac{d}{2}\right)^2 + y^2 \leqslant r^2 - \frac{d^2}{4},$$

so that it lies in a disk centered at $(d/2, 0)$ with radius h given by (1). This proves the lemma.

COROLLARY. *If a circumscribed circle exists, it is unique.*

Proof. If C_1 and C_2 were both circumscribed circles of E, they would have to have the same minimum radius r, and the corresponding disks D_1, D_2 would have nonempty intersection, since $E \subset D_1 \cap D_2$. Hence the distance d between the centers must satisfy $d < 2r$. By lemma 1, E lies in a disk of radius $h < r$, contradicting the definition of a circumscribed circle.

Proof of Theorem 2. Since E is bounded, it lies in some disk of radius M. Let

$$S = \{r \leqslant M\colon E \text{ lies in a disk of radius } r\}$$

and

$$r_0 = \inf_S r.$$

Since E contains at least two points whose distance apart is m, say, it follows that

$$r_0 \geqslant \frac{m}{2} > 0.$$

Choose a sequence of r_n in S such that

$$r_n \to r_0 \quad \text{as} \quad n \to \infty.$$

The corresponding closed disks D_n of radius r_n and center c_n, all include E. We shall show first that the points c_n tend to a limit, c_0, and second that E lies in the closed disk D_0 of radius r_0 centered at c_0.

Given $\epsilon > 0$, there exists N such that

$$r_n \leqslant r_0 + \epsilon \quad \text{for} \quad n \geqslant N.$$

For $m, n \geqslant N$, let $r = \max\{r_m, r_n\}$ and let d be the distance between c_m and c_n. Since $E \subset D_m \cap D_n$, we have

$$d < r_1 + r_2 \leqslant 2r.$$

By lemma 1, E lies in a disk D of radius h, where

$$h^2 = r^2 - \left(\frac{d}{2}\right)^2.$$

Then $h \in S$, and therefore $h \geqslant r_0$. Hence

$$\left(\frac{d}{2}\right)^2 = r^2 - h^2 \leqslant r^2 - r_0^2 = (r - r_0)(r + r_0) \leqslant 2M\epsilon.$$

If the coordinates of c_n are (x_n, y_n), it follows that for $m, n \geqslant N$,

$$\left.\begin{matrix} |x_n - x_m| \\ |y_n - y_m| \end{matrix}\right\} \leqslant d \leqslant 2\sqrt{2M\epsilon}\,.$$

Hence $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences, and must converge to x_0, y_0; so c_n converges to $c_0 = (x_0, y_0)$.

To show that the circle of radius r_0 centered at c_0 is a circumscribed circle of E, it is sufficient to show that E lies in the corresponding closed disk D_0, since E cannot lie in any disk of smaller radius, by the definition of r_0. But any point (x, y) outside D_0 satisfies

$$(x - x_0)^2 + (y - y_0)^2 = R^2, \qquad R > r_0.$$

If $\delta = R - r_0$, choose N so that $n \geqslant N$ implies

$$(x_n - x_0)^2 + (y_n - y_0)^2 < \left(\frac{\delta}{2}\right)^2$$

and

$$r_n < r_0 + \frac{\delta}{2}.$$

Then since $E \subset D_n$, by the triangle inequality every point of E has distance to c_0 at most

$$r_n + \frac{\delta}{2} < \left(r_0 + \frac{\delta}{2}\right) + \frac{\delta}{2} = R$$

so that (x, y) is not in E. This proves the theorem.

THEOREM 2. *Let E be a closed bounded set in the plane containing more than one point, and C its circumscribed circle. Then $E \cap C$ either consists of two points at the ends of a diameter of C, or else contains at least three points.*

LEMMA 2. *Under the hypotheses of theorem 2, every closed semicircle of C contains a point of E.*

Proof. Let C be the circle $x^2 + y^2 = r^2$, and let S denote the closed semi-circle: $\{(x, y) \in C: x \geqslant 0\}$. Suppose that $E \cap S$ is empty. Then, since E and S are both closed and bounded, there is a positive distance δ between them. Let C' be the circle C shifted a distance $\delta/2$ to the left. Since any point inside C and outside C' is at a distance at most $\delta/2$ from the point of S with the same y-coordinate, it follows that such points cannot be in E. Hence E lies in the intersection of the disks determined by C and C', and by lemma 1, E lies in a disk of radius less than r. But that contradicts the assumption that C is the circumscribed circle, and hence the lemma is proved.

Proof of Theorem 2. The proof is immediate from Lemma 2, since if $E \cap C$ were empty, contained a single point, or consisted of two points lying in an open semi-circle, then there would be a closed semi-circle of C containing no points of E.

Remark. The following variant of Theorem 2 also follows from Lemma 2, with a bit more argument.

THEOREM 2'. *Under the hypotheses of Theorem 2, if $E \cap C$ does not contain a pair of diametrically opposite points of C, then $E \cap C$ contains three points with the property that any open semi-circle of C contains at least one of them.*

REFERENCE

1. R. Osserman, The four-or-more vertex theorem, *Amer. Math. Monthly*, 92 (1985) 332–337.

Stirling's Series and Bernoulli Numbers

ELIAS Y. DEEBA and DENNIS M. RODRIGUEZ
Department of Applied Mathematical Sciences, University of Houston—Downtown, Houston, TX 77002

Introduction. In an article published in the MONTHLY (see [4]), Professor Victor Namias gave a simple derivation of Stirling's asymptotic series

$$n! = n^n e^{-n}\sqrt{2\pi n}\,\exp\left(\frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} + \cdots\right) \tag{1}$$

or equivalently

$$\Gamma(n) = n^{n-1}e^{-n}\sqrt{2\pi n}\,\exp\left(\sum_{k=1}^{\infty}\frac{a_k}{n^k}\right), \tag{2}$$

where the a_k's were to be determined.

Professor Namias' elegant derivation was based on the Gauss Legendre duplication formula

$$\Gamma(2n) = \frac{1}{\sqrt{\pi}}\Gamma(n)\Gamma\left(n + \frac{1}{2}\right)2^{2n-1}. \tag{3}$$

His derivation gives a recurrence relation for the coefficients a_k in the Stirling series (2). The same method applied to the triplication formula

$$\Gamma(3n) = \frac{1}{2\pi}3^{(3n-1/2)}\Gamma(n)\Gamma\left(n + \frac{1}{3}\right)\Gamma\left(n + \frac{2}{3}\right), \tag{4}$$

yields another recurrence relation for the coefficients a_k of the Stirling series. As a result he was able to derive the following two recurrence relations for the Bernoulli numbers B_k:

$$B_m = \frac{1}{2(1-2^m)}\sum_{k=0}^{m-1}2^k\binom{m}{k}B_k \tag{5}$$

$$B_m = \frac{1}{3(1-3^m)}\sum_{k=0}^{m-1}3^k(1+2^{m-k})\binom{m}{k}B_k. \tag{6}$$

Formulas (5) and (6) are based on the multiplication formulas (3) and (4), respectively, and on the well-known formula (see [1] and [3]),

$$\Gamma(n) = n^{n-1}e^{-n}\sqrt{2\pi n}\,\exp\left(\sum_{m=2}^{\infty}\frac{B_m}{m(m-1)n^{m-1}}\right). \tag{7}$$

He conjectured that an infinite number of recurrence relations for Bernoulli numbers can be obtained. He also conjectured that an infinite number of distinct recurrence relations for the a_k's can be found, all giving the coefficients of the same Stirling series (1).

In this note we present a very elementary procedure, without use of the Gauss multiplication formula, to obtain an infinite number of recurrence relations for the

Bernoulli numbers (as conjectured by Namias). In particular, we prove that

$$B_m = \frac{1}{n(1-n^m)} \sum_{k=0}^{m-1} n^k \binom{m}{k} B_k \sum_{j=1}^{n-1} j^{m-k} \tag{8}$$

is valid for any positive integer m and any positive integer $n > 1$.

Notice that (8) generalizes Namias' results, since (5) and (6) can be obtained from (8) by letting n be 2 and 3 respectively. Formula (8) was proposed as a problem in the MONTHLY (see [2]). Our proof of (8) is based only on the definition of the Bernoulli numbers as given by the formal power series expansion

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}. \tag{9}$$

Using (7) and (8) we also show that there exist an infinite number of distinct recurrence relations for the a_k's (as conjectured in [4]).

Derivation of Recurrence Relations. We shall first give the proof of the recurrence relations contained in (8).

Let n be any positive integer greater than 1. Noticing that $(1 - e^{nx})/(1 - e^x)$ is the sum of a finite geometric series, we have

$$\frac{1-e^{nx}}{1-e^x} = \sum_{j=0}^{n-1} e^{jx} = \sum_{j=0}^{n-1} \sum_{m=0}^{\infty} \frac{j^m x^m}{m!} = \sum_{m=0}^{\infty} \sum_{j=0}^{n-1} \frac{j^m x^m}{m!}.$$

Multiplying both sides by $x/(1 - e^{nx})$, we obtain

$$\frac{x}{1-e^x} = \frac{1}{n}\left(\frac{nx}{1-e^{nx}}\right) \sum_{m=0}^{\infty} \sum_{j=0}^{n-1} \frac{j^m x^m}{m!}.$$

From the definition of the Bernoulli numbers, we now have

$$\sum_{m=0}^{\infty} \frac{B_m}{m!} x^m = \left(\frac{1}{n} \sum_{m=0}^{\infty} \frac{B_m n^m x^m}{m!}\right)\left(\sum_{m=0}^{\infty} \sum_{j=0}^{n-1} \frac{j^m x^m}{m!}\right).$$

Taking the Cauchy product, we get

$$\sum_{m=0}^{\infty} \frac{B_m}{m!} xm = \frac{1}{n} \sum_{m=0}^{\infty} \sum_{k=0}^{m} \left(\frac{B_k n^k x^k}{k!} \sum_{j=0}^{n-1} \frac{j^{m-k} x^{m-k}}{(m-k)!}\right)$$

$$= \sum_{m=0}^{\infty} \left(\frac{1}{n} \sum_{k=0}^{m} \frac{B_k n^k}{k!(m-k)!} \sum_{j=0}^{n-1} j^{m-k}\right) x^m.$$

Because a power series expansion is unique, we have

$$B_m = \frac{1}{n} \sum_{k=0}^{m} n^k \binom{m}{k} B_k \sum_{j=0}^{n-1} j^{m-k}$$

$$= \frac{1}{n} \sum_{k=0}^{m-1} n^k \binom{m}{k} B_k \sum_{j=0}^{n-1} j^{m-k} + \frac{1}{n}\left(n^m B_m \sum_{j=0}^{n-1} 1\right)$$

$$= \frac{1}{n} \sum_{k=0}^{m-1} n^k \binom{m}{k} B_k \sum_{j=0}^{n-1} j^{m-k} + n^m B_m.$$

Therefore,

$$B_m = \frac{1}{n(1-n^m)} \sum_{k=0}^{m-1} n^k \binom{m}{k} B_k \sum_{j=1}^{n-1} j^{m-k} \quad \text{for all } m \geq 1,$$

which establishes (8). Next we shall exhibit distinct recurrence relations for the coefficients a_k in the Stirling series.

Equating the coefficients in (2) and (7), gives the relationship

$$a_m = \frac{B_{m+1}}{m(m+1)} \quad \text{for } m = 1, 2, 3, \ldots . \tag{10}$$

Using (8), we get

$$a_m = \frac{1}{m(m+1)} \frac{1}{n(1-n^{m+1})} \sum_{k=0}^{m} n^k \binom{m+1}{k} B_k \sum_{j=1}^{n-1} j^{m+1-k}.$$

Since $B_0 = 1$ and $B_1 = -1/2$,

$$a_m = \frac{1}{m(m+1)} \frac{1}{n(1-n^{m+1})} \left(\sum_{j=1}^{n-1} j^{m+1} - \frac{n(m+1)}{2} \sum_{j=1}^{n-1} j^m + \sum_{k=2}^{m} n^k \binom{m+1}{k} B_k \sum_{j=1}^{n-1} j^{m+1-k} \right).$$

Using (10) once again, and letting

$$f(m,n) = \sum_{j=1}^{n-1} j^{m+1} - \frac{n(m+1)}{2} \sum_{j=1}^{n-1} j^m,$$

we get

$$\begin{aligned} a_m &= \frac{1}{m(m+1)} \frac{1}{n(1-n^{m+1})} \\ &\quad \times \left(f(m,n) + \sum_{k=2}^{m} n^k \binom{m+1}{k} (k-1) k a_{k-1} \sum_{j=1}^{n-1} j^{m+1-k} \right) \\ &= \frac{1}{m(m+1)} \frac{1}{n(1-n^{m+1})} \\ &\quad \times \left(f(m,n) + \sum_{k=1}^{m-1} n^{k+1} \binom{m+1}{k+1} k(k+1) a_k \sum_{j=1}^{n-1} j^{m-k} \right). \end{aligned}$$

Hence

$$a_m = \frac{1}{(1-n^{m+1})} \left(\frac{f(m,n)}{m(m+1)n} + (m-1)! \sum_{k=1}^{m-1} \frac{n^k a_k}{(m-k)!(k-1)!} \sum_{j=1}^{n-1} j^{m-k} \right). \tag{11}$$

Notice that for each positive integer n, (11) gives a recursion formula for the sequence $\{a_m\}$. Furthermore, each recursion formula generates (1).

For example, letting $n = 2$ in (11), we get the recursion formula

$$a_m = \frac{1}{(1 - 2^{m+1})}\left(-\frac{1}{2(m+1)} + (m-1)!\sum_{k=1}^{m-1}\frac{2^k a_k}{(m-k)!(k-1)!}\right). \quad (12)$$

Letting $n = 3$ in (11) we get the recursion formula

$$a_m = \frac{1}{(1 - 3^{m+1})}\left(\frac{(1-3m)2^m - (1+3m)}{6m(m+1)} + (m-1)!\sum_{k=1}^{m-1}\frac{(2^{m-k}+1)3^k a_k}{(m-k)!(k-1)!}\right). \quad (13)$$

Formulas (12) and (13) are precisely the two recursion formulas Namias obtains in [4] using the Gauss multiplication formula.

In summary, we have obtained, in an elementary fashion, a generalization of the results in [4], and have added an infinite number of recurrent relations for the Bernoulli numbers to the ones that already exist.

REFERENCES

1. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, App. Math. Series 55, 1965.
2. E. Deeba and D. Rodriguez, solution to problem E 3237, this MONTHLY 94 (1987) 995, submitted.
3. Sir H. Jeffreys and B. Swirles (Lady Jeffreys), Methods of Mathematical Physics, Cambridge University Press, 1962.
4. Victor Namias, A simple derivation of Stirling's asymptotic series, this MONTHLY 93 (1986) 25–29.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A New Scheme for Multiple-Choice Tests in Lower-Division Mathematics

BRUCE R. JOHNSON
Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 2Y2

With the class enrollment figures for my lower-division mathematics and statistics classes rising over the hundred mark, I reluctantly made the decision to experiment with multiple-choice tests. I began with a 50-minute midterm test consisting of 20 multiple-choice questions, each in the standard five-choice format. It was a pleasure to experience the speed and accuracy of computer grading with the added bonus of question-by-question statistical analysis of class performance. I felt reasonably comfortable about the additional emphasis on error-free work due to no partial credit awards. In our high-tech age it seems that small technical errors are often harder to spot and potentially more damaging than ever before, so it probably is appropriate to put more weight on accuracy in first-year and second-year mathematics courses. Also, I found that multiple-choice provided a framework for broader, if less deep, coverage of the concepts because of its suitability for many short questions rather than a few long ones. However, I found much that troubled me more than the lack of depth. Setting the test was a slow, painful, mind-straining chore. It was very tiresome trying to anticipate, for each question, the four most common incorrect numerical answers to accompany the one correct answer. I was further troubled by the thought that some students would work backwards from the choices using a process of elimination and that guessing as a last resort would yield at least a 20% chance of success. But what bothered me most of all was the realization that the student who arrived at an answer that was not listed would be warned automatically of a mistake, while the student who arrived at a listed wrong answer would not. To warn some and not others is unfair. I rejected the possible solution of always using "none of the above" for the fifth choice, because any question having this phrase as the correct choice would reward every numerical answer except the four listed wrong answers.

But there is an approach that substantially reduces the shortcomings of standard multiple-choice tests; it is especially suitable for lower-division mathematics and statistics where most of the questions have numerical answers. For each multiple-choice question requiring a numerical answer, ten numbers are given in numerically increasing order. The correct answer might be one of the ten choices, but usually is not. The students are to select the choice nearest their computed answer. In the event of a student's answer being equidistant from the two nearest values given, the rule is to select the larger of these two choices.

Questions are much easier to set in the new format. The choices should be spaced over a feasible range at intervals narrow enough to provide a good chance

1. Find the first row, second column entry of M^{-1} where

$$M = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}.$$

(A) Does not exist (B) -1.0 (C) -0.5 (D) 0.0 (E) 0.5
(F) 1.0 (G) 1.5 (H) 2.0 (I) 2.5 (J) 3.0

$$M^{-1} = \begin{bmatrix} 0.4 & -0.6 \\ -0.2 & 0.8 \end{bmatrix}.$$

2. A research team consists of 9 men and 6 women. A project outline is to be developed by a group of 3 men and 2 women from the team. How many such groups are possible?

(A) 100 (B) 200 (C) 300 (D) 400 (E) 500
(F) 600 (G) 900 (H) 1200 (I) 2400 (J) 4800

The number of groups is $\binom{9}{3}\binom{6}{2} = 1260$.

3. A certain fabric has an average of 1.5 flaws per square meter. In a 6-square-meter piece of this fabric, what is the probability that there are no more than 11 flaws?

(A) 0.0 (B) 0.1 (C) 0.2 (D) 0.3 (E) 0.4
(F) 0.5 (G) 0.6 (H) 0.7 (I) 0.8 (J) 0.9

Since

$$\sum_{k=0}^{11} \frac{9^k}{k!} e^{-9} = 0.803,$$

the correct choice is (I).

4. Let X_1, X_2, X_3, X_4 be a random sample from a distribution with mean $\mu = 2.4$ and standard deviation $\sigma = 3.0$. Find the standard deviation of $\bar{X}$.

(A) 1.0 (B) 1.5 (C) 2.0 (D) 2.5 (E) 3.0
(F) 3.5 (G) 4.0 (H) 4.5 (I) 5.0 (J) 5.5

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{4}} = 1.5.$$

A Simple Proof of the Weierstrass Approximation Theorem

J. H. LINDSEY II
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

Let $f(x)$ be a continuous function on $I = [a, b]$. The Weierstrass theorem states that for given $\varepsilon > 0$ there exists a polynomial $p(x)$ with $|f(x) - p(x)| < \varepsilon$ on I. The general idea of the proof in this note is to split I into intervals I_i on which $f(x)$ is nearly the constant c_i, define $k_i(x)$ to be c_i on I_i and 0 elsewhere,

1. Find the first row, second column entry of M^{-1} where

$$M = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}.$$

(A) Does not exist (B) -1.0 (C) -0.5 (D) 0.0 (E) 0.5
(F) 1.0 (G) 1.5 (H) 2.0 (I) 2.5 (J) 3.0

$$M^{-1} = \begin{bmatrix} 0.4 & -0.6 \\ -0.2 & 0.8 \end{bmatrix}.$$

2. A research team consists of 9 men and 6 women. A project outline is to be developed by a group of 3 men and 2 women from the team. How many such groups are possible?

(A) 100 (B) 200 (C) 300 (D) 400 (E) 500
(F) 600 (G) 900 (H) 1200 (I) 2400 (J) 4800

The number of groups is $\binom{9}{3}\binom{6}{2} = 1260$.

3. A certain fabric has an average of 1.5 flaws per square meter. In a 6-square-meter piece of this fabric, what is the probability that there are no more than 11 flaws?

(A) 0.0 (B) 0.1 (C) 0.2 (D) 0.3 (E) 0.4
(F) 0.5 (G) 0.6 (H) 0.7 (I) 0.8 (J) 0.9

Since

$$\sum_{k=0}^{11} \frac{9^k}{k!} e^{-9} = 0.803,$$

the correct choice is (I).

4. Let X_1, X_2, X_3, X_4 be a random sample from a distribution with mean $\mu = 2.4$ and standard deviation $\sigma = 3.0$. Find the standard deviation of $\bar{X}$.

(A) 1.0 (B) 1.5 (C) 2.0 (D) 2.5 (E) 3.0
(F) 3.5 (G) 4.0 (H) 4.5 (I) 5.0 (J) 5.5

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{4}} = 1.5.$$

A Simple Proof of the Weierstrass Approximation Theorem

J. H. LINDSEY II
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

Let $f(x)$ be a continuous function on $I = [a, b]$. The Weierstrass theorem states that for given $\varepsilon > 0$ there exists a polynomial $p(x)$ with $|f(x) - p(x)| < \varepsilon$ on I. The general idea of the proof in this note is to split I into intervals I_i on which $f(x)$ is nearly the constant c_i, define $k_i(x)$ to be c_i on I_i and 0 elsewhere,

approximate $f(x)$ by the sum of the $k_i(x)$, approximate $k_i(x)$ by a normal distribution, and approximate these normal distributions by their Taylor polynomials. This gives a polynomial $f_1(x)$ approximating $f(x)$, but not closely enough. Then we similarly approximate $f(x) - f_1(x)$ by $f_2(x)$ and repeat this procedure until we have the desired $p(x)$. The details follow. The idea of successive approximations is also used by E. Schenkman [2] in his proof. The classical proof by Bernstein polynomials (i.e., binomial distributions) is given in [1].

Extend f to $J = [a - 1, b + 1]$ by defining $f(x) = f(a)$ for $x < a$ and $f(b)$ for $x > b$. Let $M = \max_J |f(x)|$. Assume $M > 0$. By uniform continuity there exists δ with $0 < \delta < 0.1$ such that $|f(x) - f(t)| < 0.01M$ if $|x - t| \le 2\delta$. Define $a(x) = e^{-(x/\delta)^2}$ and $b(x) = a(x + \delta) + a(x - \delta)$. Then for $0 \le x \le \delta$, $e^{-1} = 0 + a(0 - \delta) \le a(x + \delta) + a(x - \delta) \le a(0 + \delta) + a(\delta - \delta) = e^{-1} + 1$ (because $a(x)$ is increasing on negative real numbers and decreasing on positive real numbers). The same is true for $-\delta \le x \le 0$ since $b(x)$ is an even function. Thus $|b(x) - 1| \le 1 - e^{-1}$ for $-\delta \le x \le \delta$. Let $K = \{2i\delta: i \in \mathbb{Z} \text{ and } 2i\delta \in J\}$ and $g(x) = \sum_{y \in K} f(y)a(x - y)$. Let $x \in I$. Then $x \in [y, y + 2\delta]$ for some $y \in K$. Let $h(x) = f(y)a(x - y) + f(y + 2\delta)a(x - y - 2\delta)$. Then

$$\begin{aligned}
&|f(x)b(x - y - \delta) - h(x)| \\
&\quad = |(f(x) - f(y))a(x - y) + (f(x) - f(y + 2\delta))a(x - y - 2\delta)| \\
&\quad \le |f(x) - f(y)| + |f(x) - f(y + 2\delta)| \\
&\quad \le 0.01M + 0.01M = 0.02M
\end{aligned}$$

and

$$\begin{aligned}
|f(x) - h(x)| &\le |f(x) - f(x)b(x - y - \delta)| + |f(x)b(x - y - \delta) - h(x)| \\
&\le |f(x)|(1 - e^{-1}) + 0.02M \le M(1.02 - e^{-1}).
\end{aligned}$$

In absolute value, the contribution to $g(x)$ of the terms of $g(x)$ after the two terms of $h(x)$ is at most

$$M \sum_{i=1}^{\infty} e^{-(2i)^2} < M \sum_{i=1}^{\infty} e^{-4i} = \frac{Me^{-4}}{1 - e^{-4}} < 0.1M.$$

Allowing this $0.1M$ and another $0.1M$ for the terms before those of $h(x)$, $|f(x) - g(x)| < M(1.22 - e^{-1})$. Some number of terms in the series for $e^{(|J|/\delta)^2}$ approximates it within $0.01/|K|$. The same number of terms approximates $a(x - y)$ as closely for $y \in K$ and $x \in I$. Replacing each $a(x - y)$ by such terms converts $g(x)$ to a polynomial $f_1(x)$ with $|f(x) - f_1(x)| \le M(1.22 - e^{-1}) + |K|M(0.01)/|K| = M(1.23 - e^{-1}) < 0.9M$ for $x \in I$. Letting $f(x) - f_1(x)$ play the role of f we can find a polynomial $f_2(x)$ with $|f(x) - f_1(x) - f_2(x)| < 0.9\ (0.9M)$ for $x \in I$. Recursively, we find $f_i(x)$ with $|f(x) - \sum_{i=1}^{k} f_i(x)| < 0.9^k M$ until this is sufficiently small.

REFERENCES

1. R. R. Goldberg, Methods of Real Analysis, Blaisdell, New York, 1964.
2. E. Schenkman, The Weierstrass approximation theorem, this MONTHLY, 79 (1972) 65–67.

Statistics Within Departments of Mathematics at Liberal Arts Colleges

THOMAS L. MOORE
Department of Mathematics, Grinnell College, Grinnell, IA 50112

JEFFREY A. WITMER
Department of Mathematics, Oberlin College, Oberlin, OH 44074

In his important article "The Science of Patterns" [9], former MAA president Lynn Steen provides us with a state-of-the-discipline report on modern mathematics. Steen divides the mathematical sciences into three parts "of roughly comparable size": statistical science, core mathematics, and applied mathematics. Currently, core mathematics dominates both the faculty and curriculum of departments of mathematics at liberal arts colleges. Although Steen claims that distinctions between his three divisions may be less "intrinsic differences" than "differences in style, purpose, and history," we feel recognition of these differences can be an important benefit to the community of liberal arts colleges, where practicalities dictate that all three divisions of the mathematical sciences be housed in a single department. In particular, a broadened view of the mathematical sciences can energize a mathematics department by increasing enrollments, the number of majors, the number of students going on to graduate school, and the general level of enthusiasm for studying the mathematical sciences. Below we outline some of the essential differences between statistics and core mathematics and discuss some practical curricular implications that these differences have for a mathematics department at a liberal arts college. We concentrate on statistics rather than applied mathematics because we are statisticians.

Statistics is underrepresented among mathematics faculty at liberal arts colleges in the United States. A recent survey [8] of mathematics departments at liberal arts colleges suggests that approximately half of all such departments have no one with an advanced degree in statistics and only 12.5% have more than one such person. And, unlike the situation at many universities, if statisticians are employed at liberal arts colleges they will generally be housed in the department of mathematics, since mathematics is the traditional liberal arts discipline most closely aligned with statistics. Because the responsibility of statistics education at most liberal arts colleges rests with the mathematics department, it is imperative that the department recognize the fundamental differences between statistics and core mathematics and ensure that their statistics curriculum reflects these differences.

Core mathematics and statistics differ in two fundamental ways. Both fields look for structure and patterns, but core mathematics looks in the abstract arena of space and number while statistics looks at data from other, nonmathematical subject areas. Hence core mathematics and statistics differ in their *objects* of study.

They also differ in their methods. The emphasis in mathematical thinking is on deduction. For example, the axioms of classical group theory form a framework from which theorems can be derived. Naturally, in the discovery of new truths inductive reasoning and exploration are essential tools. In developing a new

theorem in group theory one may observe patterns in known groups with certain characteristics. If all such groups examined exhibit a certain property then a conjecture may emerge and a proof be sought. Nevertheless, in this endeavor the emphasis is on the deduction that establishes the theorem.

Statistical thinking, on the other hand, focuses on a dialog between models and data. For example, suppose one wants to develop a predictive relationship between a student's SAT score and his or her college GPA, using linear regression. One can postulate a simple linear relationship with independent, normally distributed deviations from the line and from this model infer many features of the population from which the data came. For example, one could construct an interval estimate of the GPA of a student with an 1150 SAT. However, the linear regression model is *only* a model of reality. All statistical models are tentative. If diagnostic plots show that the model does not fit the data, then a more complicated model will have to be considered, the inferences redrawn, and the new model scrutinized for appropriateness in its own right. This dialog between model and data is a fundamental feature of statistical thinking.

A critical aspect of this dialog between model and data is the quality of the data. For example, by properly choosing the sample or by designing a good experiment the statistician can usually simplify analysis and strengthen the interpretability of the model. It is therefore very important to present these ideas when teaching statistics. Even an introductory course should give attention to ideas such as sampling and experimental bias, the role of randomization, the idea of pairing observations, etc. Unfortunately many introductory courses and textbooks ignore these concepts.

Current trends in statistical research reflect the fundamental difference between statistics and core mathematics. Consider first the area of regression diagnostics. Often in linear regression one or more points may be anomalous. The fitted model may not fit these points well (i.e., they may be outliers) or these points may have enormous influence on the fitting process itself. Much recent work has gone into developing diagnostic procedures for identifying outliers and influential points in complicated regression problems (see, e.g., [3]). These procedures play a key role in the iterative cycle of data analysis in which a statistician fits a model, then checks the fit using diagnostic methods, then if necessary fits a modified model, and so on.

A typical regression analysis using diagnostics will depend heavily on the computer. Nevertheless, other areas of modern statistical research make even greater computing demands. Classical statistical inference rests upon a precise model specification. For example, in the SAT/GPA example above we need both the underlying straight-line and the normally distributed errors to create the interval estimate. The *bootstrap* is a method for making a statistical inference without standard model specifications. Using the bootstrap one repeatedly draws pseudo-samples from the data at hand and uses these pseudo-samples as the basis of, for example, an interval estimate of a parameter. The bootstrap is particularly useful in situations where the model-based theory is poorly developed. A good survey of this technique is given by Efron and Tibshirani [4].

The area of dynamic graphical data analysis provides a third example of statistics research that is quite distinct from core mathematics. The thrust of this

research is using modern computer graphics to develop methods that allow the data analyst to discover the structure in complex, multidimensional data sets. Becker, Cleveland, and Wilks [1] give a good survey of this area.

For each of the past few summers, a small group of statisticians from liberal arts colleges has met to discuss the nature of statistics within liberal arts colleges. These Statistics in the Liberal Arts Workshops (SLAW) have considered three aspects of statistics within the liberal arts setting: the teaching of statistics, the role of statistics within the mathematics major, and the role of a statistician as a general campus resource (see [8]). A key conclusion was that a vibrant statistics curriculum includes real data and applied statistics and that every mathematics department should offer at least one data-driven statistics course that counts for mathematics major credit and that preferably can be taken early in a student's career. Several models are possible for such a course: (1) Make an existing introductory course data-driven and count it for the major. (2) Add a (one- or two-credit) supplement to the traditional mathematical statistics course. (3) Teach new and different applied and data-driven statistics courses at the introductory level for mathematics majors.

Model 1 may be the easiest to implement. Most liberal arts college mathematics departments teach a course in introductory statistics. Why not allow credit for the mathematics major for such a course? The traditional response to this question is that such a course is not mathematical enough, where "mathematical" is taken to mean "core mathematical." However, once statistics is given equal status as a mathematical science, this objection will vanish *if* such introductory courses are taught with *statistical rigor*, that is, if they are data-driven explorations of the discipline of statistics, as opposed to the dry presentations of formulas that students often see. Model 1 together with the traditional mathematical statistics course then becomes similar to the model most of us now use for teaching calculus: an introduction that teaches basic techniques and the flavor of the subject followed later by a theoretical real analysis course.

It is natural for mathematicians teaching statistics to emphasize the mathematical (i.e., theoretical, deductive) side of the discipline. Such an approach makes things easier for the teacher, but gives only a partial picture of statistics. For within such a curriculum students who might be drawn to statistical thinking see little of it and potential mathematical scientists (and potential mathematics majors!) are lost.

A good data-driven course uses *real* data to teach statistical thinking. Experience suggests that real data are a far more powerful motivating force than are artificial data. Statisticians—and their students—are interested in understanding and solving real problems. The analysis of data should be the focus of an introductory course. Students should discuss the real-world problems the data were collected to solve, the quality of the data, and their effective analysis. Exercises should be built around real data sets and students might even design projects that require them to produce and analyze their own data. The emphasis of such a course is on understanding the broad concepts that apply to most statistical problem solving.

At least some portion of the course should deal with larger data sets and their analysis using a statistical computer package. It is important to include large data sets and computers for several reasons. First, a good interactive package is a basic

tool for the modern data analyst and it is good for the student to learn about one. Second, most real data sets are too large for hand calculations. Third, larger data sets can better motivate some techniques than smaller ones. For example, finding outliers in a small data set can be done by scanning a listing of the data, but with a larger data set one turns naturally to statistical graphics.

The widespread availability of computers and easy-to-use statistics packages (e.g., Minitab, as opposed to SAS or SPSS-X, which are powerful but not user-friendly) makes it possible to avoid tedious calculation when analyzing real data and, instead, to concentrate on the relevant statistical concepts. When the computer is used, students can spend less time on calculation and more time on the important task of learning what the numbers mean. This in turn gives the teacher more room to test understanding of concepts. Dealing with statistical concepts can be difficult, but students should not be able to avoid thinking by hiding behind formulas (and neither should their teachers). We have found, after a period of adjustment, that de-emphasizing formulas helps us think more clearly about the concepts we are trying to teach.

Teaching a data-driven course can be a challenge. Such a course will require a different style of teaching for most faculty, with more discussion and less lecture. More time will be spent on the computer with a statistical package and, possibly, simulations. Office time will be spent helping students work out the very nonmathematical aspects of the design and analysis of their own projects. Teaching a data-driven course will be difficult the first time, but the incorporation of the preceding ideas can be approached in steps and can result in a very rewarding teaching experience.

A good textbook that teaches statistical thinking can be of great help. Three good examples, aimed at progressively more sophisticated audiences, are [**6**], [**5**], and [**7**]. A recent article by Cobb [**2**] provides excellent guidelines on what to look for in a textbook and also reviews several. We present a short list of resources and readings in the appendix to aid those who want to learn more about incorporating real examples and applications into statistics courses. A longer bibliography can be found in [**8**].

Model 2, the supplement to the mathematical statistics, is currently being used at Oberlin College, where the two-semester, six-credit sequence in probability and mathematical statistics is augmented with a one-credit supplement offered during the second semester. The course, called Data Analysis, meets once per week and covers various topics from applied statistics, including the tools of exploratory data analysis, normal probability plots and transformations of data, control charts, applied linear regression, and analysis of variance. All of the topics are introduced with computer applications to real data. In the spring of 1989 the course centered on an open-ended class project of analyzing a set of over 200 variables from a survey of college librarians at 97 liberal arts colleges, so that the students could gain first-hand experience as statisticians working on a real problem.

Several implementations of model 3—new, applied, and introductory level courses for mathematics majors—currently exist and surely others are possible. Mt. Holyoke College teaches an applied regression course that carries a calculus and linear algebra prerequisite. The students are required to do individual projects in which they find or produce their own data. They must write a final report and present their project findings to the class. St. Lawrence University offers an

applied time series and forecasting course with a year of calculus as a prerequisite. This course uses much real data that are analyzed using a computer package. Both Carleton College and Swarthmore College offer courses in applied multivariate analysis that require linear algebra as a prerequisite. Although few theorems, if any, are proved in these courses, they all require some mathematical sophistication of their audiences so that proper understanding of rather elaborate statistical models is possible.

When statistics secures a greater role in the mathematics department of a liberal arts college, everyone gains. The department becomes more properly a department of mathematical sciences as described by Steen. Students more inclined toward statistics than core mathematics have an additional point of entry into the mathematics major and the college may send more students on to careers in the mathematical sciences. At the same time, these benefits require that the differences between statistics and core mathematics be appreciated and allowed for. Ideally, a department will hire a statistician (or two) to develop a solid statistics curriculum. Short of that, mathematicians presently teaching statistics should develop a data component in their statistics curriculum along the lines outlined above. A data-driven, concept-oriented approach can breathe new life into your statistics courses.

We gratefully acknowledge the help of our fellow SLAW participants in formulating the ideas in this paper: Donald L. Bentley of Pomona, George W. Cobb of Mt. Holyoke, Homer T. (Pete) Hayslett, Jr. of Colby, Gudmund Iversen of Swarthmore, Anju Joglekar of Smith, Robin H. Lock of St. Lawrence, David S. Moore of Purdue, Rosemary Roberts of Bowdoin, and Frank Wolf of Carleton.

We also thank the referee for helpful comments.

Appendix

The following are introductory textbooks with a special strength or focus. Even experienced teachers should find them interesting.

Items 1–6 list textbooks at the elementary level.

1. J. Devore and R. Peck, *The Exploration and Analysis of Data*, 1986, West Publishing, St. Paul.
 This introductory textbook is loaded with real data from all areas of application.
2. L. Koopmans, *Introduction to Contemporary Statistical Methods*, 2nd ed., 1987, Duxbury, Boston.
 Koopmans was among the first authors to thoroughly integrate the modern topics of exploratory data analysis, graphics, and robust methods into a coherent textbook.
3. B. F. Ryan, B. L. Joiner, and T. A. Ryan, Jr., *Minitab Handbook*, 2nd ed., 1985, PWS-Kent, Boston.
 The Handbook is an excellent companion textbook to the highly popular statistical package Minitab. The book discusses the several larger data sets that come with the package and includes many good exercises about them.
4. D. S. Moore, *Statistics: Concepts and Controversies*, 1989, W. H. Freeman, New York.
5. D. Freedman, R. Pisani, and R. Purves, *Statistics*, 1978, W. W. Norton, New York.

6. D. S. Moore and George McCabe, *Introduction to the Practice of Statistics*, 1989, W. H. Freeman, New York.
 We referred to these three in the article.

Items 7 and 8 are textbooks for more advanced students.

7. J. A. Rice, *Mathematical Statistics and Data Analysis*, 1988, Wadsworth, Belmont, Calif.
 This introductory textbook aimed at advanced undergraduates provides that audience with a true appreciation of the range of statistics and features the use of real data, exploratory data analysis and graphics, and a concern for model assumptions.
8. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, 1978, Wiley, New York.
 This textbook provides a lucid, conceptual introduction to statistics and experimentation for an audience of engineers, chemists, and other industrial scientists.

Items 9 and 10 are excellent supplemental reading for students and teachers.

9. W. S. Peters, *Counting for Something: Statistical Principles and Personalities*, 1987, Springer-Verlag, New York.
 This charming and casual introduction to elementary statistical concepts gives the reader an excellent sense of the history of the discipline.
10. J. M. Tanur, F. Mosteller, W. H. Kruskal, E. L. Lehmann, R. F. Link, R. S. Pieters, and G. R. Rising, *Statistics: A Guide to the Unknown*, 3rd ed., 1989, Wadsworth, Belmont, Calif.
 This is an anthology of short, nontechnical articles on a wide variety of statistical applications.

REFERENCES

1. Richard A. Becker, William S. Cleveland, and Allan R. Wilks, Dynamic graphics for data analysis, *Statistical Science*, 1 (1986) 355–395.
2. George W. Cobb, Introductory textbooks: a framework for evaluation, *Journal of the American Statistical Association*, 82 (1987) 321–339.
3. R. Dennis Cook and Sanford Weisberg, Residuals and Influence in Regression, 1982, Chapman and Hall, London.
4. B. Efron and R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, 1 (1986) 54–77.
5. David Freedman, Robert Pisani, and Roger Purves, Statistics, 1978, W. W. Norton, New York.
6. David S. Moore, Statistics: Concepts and Controversies, 2nd ed., 1985, W. H. Freeman, New York.
7. David S. Moore and George P. McCabe, Introduction to the Practice of Statistics, 1989, W. H. Freeman, New York.
8. Thomas L. Moore and Rosemary A. Roberts, Statistics at liberal arts colleges, *American Statistician*, 43 (1989) 80–85.
9. Lynn A. Steen, The science of patterns, *Science*, 240 (1988) 611–616.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, MICHAEL A. FILASETA, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before October 31, 1991 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3439. *Proposed by Jane Friedman, Widener University, Chester, PA.*

If M and N are nonnegative integers, prove that

$$\binom{M+N}{M} = \sum_{0 \leq a \leq (M-1)/2} \binom{M-a-1}{a}\binom{N+a}{2a+1} + \sum_{0 \leq a \leq M/2} \binom{M-a}{a}\binom{N+a}{2a}.$$

E 3440. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Let A be a 3 by 3 magic matrix with real elements; i.e., there is a nonzero real number s such that each row of A sums to s, each column of A sums to s, the main diagonal of A sums to s, and the counter-diagonal of A sums to s.

(i) Show that if A is also nonsingular, then A^{-1} is magic.

(ii) Show that A has the form

$$\begin{pmatrix} s/3+u & s/3-u+v & s/3-v \\ s/3-u-v & s/3 & s/3+u+v \\ s/3+v & s/3+u-v & s/3-u \end{pmatrix},$$

where u and v are arbitrary, and is nonsingular if and only if $v^2 \neq u^2$.

E 3441. *Proposed by Xu Chenglong, Shanghai University of Science and Technology, China.*

Suppose $0 \leqslant p < 1$ and $q \geqslant 0$. Put

$$U_n = \sum_{k=0}^{n} \binom{n}{k} p^{n-k} q^k \frac{1}{k!}.$$

Determine $\lim_{n\to\infty} U_n$.

E 3442. *Proposed by Ray Wylie, Furman University, Greenville, SC.*

Given a sequence $\{b_n\}_{n=0}^{\infty}$ of real numbers such that $b_n = 0$ for n sufficiently large, put $B_s = \sum_{t=0}^{\infty} b_{mt+s}$ $(s = 0, 1, \ldots, m-1)$, and let us say that the sequence $\{b_n\}_{n=0}^{\infty}$ has property P_m, where m is a positive integer, if

$$B_0 = B_1 = \cdots = B_{m-1}.$$

Suppose $a_1, a_2, \ldots, a_r$ are positive integers not necessarily distinct, and let $C(n)$ be the number of r-tuples $(n_1, n_2, \ldots, n_r)$ of integers such that

$$n = n_1 + n_2 + \cdots + n_r, \quad 0 \leqslant n_i \leqslant a_i \quad \text{for } i = 1, 2, \ldots, r.$$

Prove that the k sequences

$$\{C(n)\}_{n=0}^{\infty}, \{nC(n)\}_{n=0}^{\infty}, \ldots, \{n^{k-1}C(n)\}_{n=0}^{\infty}$$

all have property P_m if and only if at least k of the integers $a_1, a_2, \ldots, a_r$ are congruent to -1 modulo m.

E 3443. *Proposed by Călin Popescu, St. Michiels Brugge, Belgium.*

Let A_i, $i = 0, 1, \ldots, 5$, denote the vertices of a hexagon inscribed in a circle and let B_i denote the intersection of the straight lines A_iA_{i+2} and $A_{i+1}A_{i+3}$, for $i = 0, 1, \ldots, 5$, the indices being computed modulo 6. Prove that, if the triangles $A_0A_2A_4$ and $A_1A_3A_5$ have the same orthocenter, then the straight lines B_iB_{i+3}, $i = 0, 1, 2$, are concurrent. (The orthocenter of a triangle is the intersection of its three altitudes.)

E 3444. *Proposed by Wilbur Jónsson, McGill University, Montreal, Canada and Gérard Letac, Université Paul Sabatier, Toulouse, France.*

Let E be a finite-dimensional vector space over a field K and let $L(E)$ be the space of endomorphisms of E. Suppose L_1 and L_2 are subspaces of $L(E)$ such that $L(E) = L_1 + L_2$ and $xy + yx = 0$ for all (x, y) in $L_1 \times L_2$.

(i) If $\mathrm{char}(K) \neq 2$, prove that either $L_1 = 0$ or $L_2 = 0$.

(ii) If $\mathrm{char}(K) = 2$, show that the conclusion of (i) need not hold.

Editorial comment. We remark that trivially

$$R = S = T = \mathbb{Q} \qquad \text{when } n = 1, 2,$$

$$R = \mathbb{Q},\ S = T = \mathbb{Q}(\sqrt{3}) \qquad \text{when } n = 3,$$

$$R = S = \mathbb{Q},\ T \text{ is undefined} \qquad \text{when } n = 4.$$

Several solvers noted that the problem is partially solved in D. H. Lehmer, "A Note on Trigonometric Algebraic Numbers," this MONTHLY 40(1938) 165–166 and in Ivan Niven, *Irrational Numbers*, Carus Monograph No. 11, pp. 37–41.

Solved also by S. F. Barger, D. Callan, M. R. Darafsheh (Iran), H. M. W. Edgar, D. W. Koster, D. E. Manes, R. A. Mena, National Security Agency Problems Group, S. Preis, J. Spencer, R. Stong, S. V. Ullom, and the proposer.

A Threshold for Divergence?

E 3354 [1989, 838]. *Proposed by V. A. Alexandrov and N. S. Dairbekov, Institute of Mathematics, Novosibirsk, USSR.*

Suppose M: $[1, \infty) \to (e, \infty)$ is a nondecreasing function such that

$$\int_1^\infty \frac{dt}{M(t)} = \infty. \tag{*}$$

(a) Prove that

$$\int_1^\infty \frac{dt}{t \log M(t)} = \infty.$$

(b) Show that there exist nondecreasing functions M satisfying (*) such that

$$\int_1^\infty \frac{dt}{t \log M(t) \log\log M(t)} < \infty.$$

Solution by Kenneth Schilling, University of Michigan, Flint. For (a) suppose that $\int_1^\infty dt/\{t \log M(t)\} < \infty$. By the substitution $u = \log t$ we obtain $\int_0^\infty du/\log M(e^u) < \infty$. Since $1/\log M(e^u)$ is nonincreasing, we have

$$(v/2)/\log M(e^v) \leqslant \int_{v/2}^{v} du/\log M(e^u) \to 0$$

as $v \to \infty$. This implies $v/\log M(e^v) < 1/2$ for large v, which is equivalent to $e^{2v} < M(e^v)$ or $e^v/M(e^v) < e^{-v}$ for large v. Hence, $\int_0^\infty e^v\, dv/M(e^v) < \infty$, which yields $\int_0^\infty dt/M(t) < \infty$ via the substitution $t = e^v$. Thus (a) is proved.

For (b) we construct an example by letting $k_n = \exp\{\exp[\exp(n)]\}$ for $n = 0, 1, 2, \ldots$ and putting $M(t) = k_n$ for $k_{n-1} \leqslant t < k_n$ for $n = 1, 2, \ldots$. Then

$$\int_{e^e}^\infty dt/M(t) = \sum_{n=1}^{\infty} k_n^{-1}(k_n - k_{n-1}) = \infty,$$

Editorial comment. We remark that trivially

$$R = S = T = \mathbb{Q} \qquad \text{when } n = 1, 2,$$

$$R = \mathbb{Q},\ S = T = \mathbb{Q}(\sqrt{3}) \qquad \text{when } n = 3,$$

$$R = S = \mathbb{Q},\ T \text{ is undefined} \qquad \text{when } n = 4.$$

Several solvers noted that the problem is partially solved in D. H. Lehmer, "A Note on Trigonometric Algebraic Numbers," this MONTHLY 40(1938) 165–166 and in Ivan Niven, *Irrational Numbers*, Carus Monograph No. 11, pp. 37–41.

Solved also by S. F. Barger, D. Callan, M. R. Darafsheh (Iran), H. M. W. Edgar, D. W. Koster, D. E. Manes, R. A. Mena, National Security Agency Problems Group, S. Preis, J. Spencer, R. Stong, S. V. Ullom, and the proposer.

A Threshold for Divergence?

E 3354 [1989, 838]. *Proposed by V. A. Alexandrov and N. S. Dairbekov, Institute of Mathematics, Novosibirsk, USSR.*

Suppose M: $[1, \infty) \to (e, \infty)$ is a nondecreasing function such that

$$\int_1^\infty \frac{dt}{M(t)} = \infty. \tag{*}$$

(a) Prove that

$$\int_1^\infty \frac{dt}{t \log M(t)} = \infty.$$

(b) Show that there exist nondecreasing functions M satisfying (*) such that

$$\int_1^\infty \frac{dt}{t \log M(t) \log\log M(t)} < \infty.$$

Solution by Kenneth Schilling, University of Michigan, Flint. For (a) suppose that $\int_1^\infty dt/\{t \log M(t)\} < \infty$. By the substitution $u = \log t$ we obtain $\int_0^\infty du/\log M(e^u) < \infty$. Since $1/\log M(e^u)$ is nonincreasing, we have

$$(v/2)/\log M(e^v) \leqslant \int_{v/2}^{v} du/\log M(e^u) \to 0$$

as $v \to \infty$. This implies $v/\log M(e^v) < 1/2$ for large v, which is equivalent to $e^{2v} < M(e^v)$ or $e^v/M(e^v) < e^{-v}$ for large v. Hence, $\int_0^\infty e^v\, dv/M(e^v) < \infty$, which yields $\int_0^\infty dt/M(t) < \infty$ via the substitution $t = e^v$. Thus (a) is proved.

For (b) we construct an example by letting $k_n = \exp\{\exp[\exp(n)]\}$ for $n = 0, 1, 2, \ldots$ and putting $M(t) = k_n$ for $k_{n-1} \leqslant t < k_n$ for $n = 1, 2, \ldots$. Then

$$\int_{e^e}^\infty dt/M(t) = \sum_{n=1}^{\infty} k_n^{-1}(k_n - k_{n-1}) = \infty,$$

since $k_{n-1}/k_n \to 0$ as $n \to \infty$. On the other hand,

$$\int_{e^e}^{\infty} dt/\{t \log M(t) \log\log M(t)\} = \sum_{n=1}^{\infty} \int_{k_{n-1}}^{k_n} dt/\{te^{e^n}e^n\}$$

$$= \sum_{n=1}^{\infty} (e^{e^n} - e^{e^{n-1}})/(e^{e^n}e^n) < \sum_{n=1}^{\infty} 1/e^n < \infty.$$

Solved also by L. Shepp, R. Stong, W.-B. Zhang, and a referee. Part (a) was solved by M. Falkowitz (Israel) and the proposer. One incorrect solution was received.

The Limit of a Dirichlet Integral

E 3357 [1989, 927]. *Proposed by N. I. Alexandrova, Institute of Mining, Novosibirsk, USSR.*

Evaluate

$$\lim_{y \to +\infty} \int_0^{+\infty} \frac{\sin x^2}{x^3} \sin(xy)\, dx.$$

Solution by K. F. Andersen, University of Alberta, Edmonton, Canada. The value of the limit is $\pi/2$. To see this define a function f on $(0, \infty)$ by putting

$$f(x) = \frac{\sin x^2}{x^3} - \frac{1}{x} \quad \text{for } 0 < x < 1, \qquad f(x) = \frac{\sin x^2}{x^3} \quad \text{for } x \geqslant 1.$$

Since $-x^3/6 < (\sin x^2)/x^3 - 1/x < 0$ for $0 < x < 1$, we see that $f \in L(0, \infty)$. Hence, we may write

$$\int_0^{\infty} \frac{\sin x^2}{x^3} \sin(xy)\, dx = \int_0^{\infty} f(x) \sin(xy)\, dx + \int_0^1 \sin(xy) \frac{dx}{x}.$$

The first term on the right tends to zero as $y \to \infty$ by the Riemann–Lebesgue lemma, while the second term equals

$$\int_0^y (\sin t) \frac{dt}{t},$$

which, as is well known, tends to $\pi/2$ as $y \to \infty$.

Editorial comment. Allen Pedersen and Kee-Wai Lau used arguments like that above to reduce the problem to other known integrals. Such integrals (and the integral $\int_0^{\infty} t^{-1}(\sin t)\, dt$) are often evaluated via the Cauchy residue theorem; however, Michael Fichter was the only solver who used the residue theorem directly.

Klaus Schürger and the Chico Problem Group both obtained the generalization

$$\lim_{y \to \infty} \int_0^{\infty} x^{-(1+r)} \sin x^r \sin(xy)\, dx = \pi/2$$

for any fixed positive number r.

Several solvers calculated the limit to be 5.4415643575.... The early terms of the sequence are $x_1 = 3$, $x_2 = 4.754887\ldots$, $x_3 = 5.419336\ldots$. D. B. Tyler remarked that x_4 and the limit agree to about 15 decimal places, but that x_5 and the limit agree to over 10^{12} decimal places. The latter fact follows from the inequality $(**)$ of solution II if we replace 5 by 5.4.

34 correct solutions, 5 incorrect solutions, and one incomplete solution were received.

Falsity of a Sharper Form of Carmichael's Conjecture

E 3361 [1990, 63]. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Prove or give a counterexample to the following conjecture concerning the Euler totient function: For every even positive integer n there is an odd positive integer k such that

$$\varphi(k) = \varphi(n).$$

Solution I by L. L. Foster, California State University, Northridge. We show that the equation $\phi(x) = 2^{16} \cdot 257$ has even solutions but no odd solutions. The even solutions are

$$x = 2^{9-\varepsilon_0-2\varepsilon_1-4\varepsilon_2} \cdot 3^{\varepsilon_0} \cdot 5^{\varepsilon_1} \cdot 17^{\varepsilon_2} \cdot 257^2,$$

where each ε_i is 0 or 1. Suppose there were an odd number k such that $\phi(k) = 2^{16} \cdot 257$. If p is a prime factor of k, then either $p - 1 = 2^j$ for some j with $1 \leqslant j \leqslant 16$ or $p - 1 = 2^j \cdot 257$ for some j with $1 \leqslant j \leqslant 16$. But $2^j \cdot 257 + 1$ is composite for $j = 1, 2, \ldots, 16$ (being divisible by 67 when $j = 7$, divisible by 1123 when $j = 15$, and divisible by 3, 5, 7, or 17 for the other values of j). Thus the only possible prime factors of k are $3, 5, 17, 257, 65537$, with the prime 257 occurring exactly twice and any other of these primes occurring at most once. But clearly 65537 cannot occur in conjunction with 257 and, if ε_0, ε_1, and ε_2 are each 0 or 1, then

$$\phi(3^{\varepsilon_0} \cdot 5^{\varepsilon_1} \cdot 17^{\varepsilon_2} \cdot 257^2) = 2^{8+\varepsilon_0+2\varepsilon_1+4\varepsilon_2} \cdot 257 < 2^{16} \cdot 257.$$

Thus there is no odd k such that $\phi(k) = 2^{16} \cdot 257 = \phi(2^9 \cdot 257^2)$, so that $n = 2^9 \cdot 257^2 = 33817088$ is a counterexample to the conjecture of the problem. It is not difficult to show that 33817088 is in fact the minimal counterexample; this can also be verified by computer.

Solution II by R. J. Simpson, Curtin University, Perth, Australia. We shall prove the following rule. A necessary and sufficient condition for the equation $\phi(x) = 2^m = \phi(2^{m+1})$ to have an odd solution x is that any 2^j occurring in the binary expansion of m must be such that $2^{2^j} + 1$ is prime. To prove this rule suppose that k is odd and that $\phi(k) = 2^m$. Then k must be the product of distinct primes of the form $2^s + 1$. Now a necessary (but not sufficient) condition for $2^s + 1$ to be prime is that s be a power of 2. Thus we must have

$$k = \prod_{i=1}^{r} (2^{2^{j_i}} + 1), \qquad 0 \leqslant j_1 < j_2 < \cdots < j_r,$$

where each factor in the product is a prime. But then

$$2^m = \phi(k) = \prod_{i=1}^{r} 2^{2^{j_i}},$$

or

$$m = \sum_{i=1}^{r} 2^{j_i},$$

and the necessity is proved. The sufficiency is immediate by the same argument.

In particular, since $2^{32} + 1 = 641 \cdot 6700417$, it follows that if the binary expansion of m contains 2^5, then $\phi(x) = 2^m = \phi(2^{m+1})$ has no odd solution x.

Editorial comment. Most solvers settled the problem by observing that the equation $\phi(x) = 2^{32}$ has even solutions, namely,

$$x = 2^{33-\varepsilon_0-2\varepsilon_1-4\varepsilon_2-8\varepsilon_3-16\varepsilon_4} \cdot 3^{\varepsilon_0} \cdot 5^{\varepsilon_1} \cdot 17^{\varepsilon_2} \cdot 257^{\varepsilon_3} \cdot 65537^{\varepsilon_4},$$

where each ε_i is 0 or 1, but that (by an argument similar to that of Solution I) the equation $\phi(x) = 2^{32}$ has no odd solutions. By contrast, if $0 \leqslant m \leqslant 31$ and the binary expansion of m is given by

$$m = \delta_0 + 2\delta_1 + 4\delta_2 + 8\delta_3 + 16\delta_4,$$

where each δ_i is 0 or 1, then $\phi(x) = 2^m$ has the odd solution

$$x = 3^{\delta_0} \cdot 5^{\delta_1} \cdot 17^{\delta_2} \cdot 257^{\delta_3} \cdot 65537^{\delta_4}.$$

Using the fact that $2^{2^j} + 1$ is composite for $5 \leqslant j \leqslant 21$, Eugene E. Ilgenfritz, David W. Doster, and James G. Marickel each noted that $\phi(x) = 2^m$ has no odd solutions for $32 \leqslant m \leqslant 2^{22} - 1 = 4194303$. (This assertion also follows from the rule of Solution II.)

A counterexample to the conjecture of the problem is given in Exercise 14 on page 234 of W. Sierpiński's *Elementary Theory of Numbers*, Monografie Mat. Volume 42, Warsaw, 1964. Sierpiński essentially proved that if $m \equiv 32 (\mathrm{mod}\ 64)$, then there are no odd solutions x of $\phi(x) = 2^m = \phi(2^{m+1})$. Again this is a special case of the rule of Solution II.

No other solver found the minimal counterexample given by Foster in Solution I. Richard Stong, Douglas B. Tyler, and Jack V. Wales, Jr. each gave a solution similar to Solution II.

Several solvers remarked that the conjecture mentioned in the problem is a stronger form of Carmichael's conjecture that the equation $\phi(x) = C$ never has exactly one solution, but that the demonstrable falsity of this conjecture does not really help to decide Carmichael's conjecture.

Solved also by 63 other readers and the proposer.

An Equation in Towers of Powers

E 3364 [1990, 64]. *Proposed by H. W. Lenstra, University of California, Berkeley.*

Determine whether there exist positive integers a, b, n, m such that $a \neq b$, $n \geqslant 2$, $m \geqslant 2$, and

$$a^{a^{\cdot^{\cdot^{a}}}} \Big\} n = b^{b^{\cdot^{\cdot^{b}}}} \Big\} m.$$

Solution by James Duemmel and Allen Mauney, Western Washington University, Bellingham, WA. There is no solution. Suppose there is a solution; we may assume $1 < a < b$. If the relation is true, then we have an equation of the form $a^x = b^y$. Here a and b must have exactly the same prime divisors. Let $a = \prod_{i=1}^{t} p_i^{\alpha_i}$ and $b = \prod_{i=1}^{t} p_i^{\beta_i}$, with all α_i and β_i positive and $\alpha_i x = \beta_i y$ for all i. Letting $z = x/y$, we have $\beta_i = z\alpha_i$ for all i, and hence $b = a^z$.

Next we show that z must be an integer. Since $x = a^u$ and $y = b^v$ for some positive integers u and v, we have $z = \prod_{i=1}^{t} p_i^{(\alpha_i u - \beta_i v)}$. These exponents are integers, and if they are all positive z will be an integer. From $a^x = b^y$, we have $z = \log b/\log a$. Also $x > y$ since $a < b$, and hence $a^u > b^v$. This implies $u/v > \log b/\log a = z = \beta_i/\alpha_i$. Hence $\alpha_i u - \beta_i v > 0$ for all i.

From the expression for z, we have $z = a^w$, where $w = u - zv > 0$. We also have $u = a^s$ for some positive integer s, since $a < b$ implies that $n \geqslant 3$. Substituting for z in the expression for w, we obtain $w + a^w v = u = a^s$. Since $v \geqslant 1$, the left side is larger than a^w. Hence $s \geqslant w$ and a^w divides u. Hence a^w must also divide w. With $a > 1$, this is not possible.

Editorial comment. Edward B. Burger and Peter D. Kohn pointed out that the equation has solutions when a and b are not integers. For example, $(1/2)^{1/2} = (1/4)^{1/4}$.

Solved also by B. Beckmann, E. B. Burger & P. D. Kohn (students), D. Callan, K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), J. McHugh, J. G. Merickel (student), J.-M. Monier (France), T. S. Norfolk, E. Preissman (Switzerland), R. Stong, J. S. Sumner & K. L. Dove & E. Toro, W.-B. Zhang, Central Michigan University Problem Group, and the proposer.

ADVANCED PROBLEMS

6658. *Proposed by L. Van Hamme, Free University of Brussels, Belgium.*

Define a sequence of integers by

$$a(0) = 1, \quad a(n) = \sum_{r=0}^{n-1} \binom{n}{r} a(r) \qquad \text{for } n \geqslant 1,$$

so that $\sum_{n=0}^{\infty} a(n)x^n/n! = (2 - e^x)^{-1}$ for $|x| < \log 2$. (This is sequence 1191 in N. J. Sloane's *Handbook of Integer Sequences*, New York, Academic Press, 1973.)

Prove that if p is a prime number and m is an integer not divisible by p, then

$$a(mp^k + s) \equiv a(mp^{k-1} + s) \pmod{p^k}$$

for k a positive integer and s a nonnegative integer.

6659. *Proposed by Abraham Ungar, North Dakota State University, Fargo.*

Let $\mathbb{R}_c^3$ be the subset of the Euclidean 3-space $\mathbb{R}^3$ given by the equation

$$\mathbb{R}_c^3 = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| < c\},$$

where c is a positive constant. In the special theory of relativity c represents the speed of light, and the elements $\mathbf{x}$ of $\mathbb{R}_c^3$ are *admissible velocities*. The relativistic

velocity composition law is given by the equation

$$\mathbf{x}*\mathbf{y} = \frac{\mathbf{x}+\mathbf{y}}{1+\mathbf{x}\cdot\mathbf{y}/c^2} + \frac{1}{c^2}\cdot\frac{\gamma_{\mathbf{x}}}{\gamma_{\mathbf{x}}+1}\cdot\frac{\mathbf{x}\times(\mathbf{x}\times\mathbf{y})}{1+\mathbf{x}\cdot\mathbf{y}/c^2}, \qquad \mathbf{x},\mathbf{y}\in\mathbb{R}_c^3,$$

where $\gamma_{\mathbf{x}}$ is the *Lorentz factor*

$$\gamma_{\mathbf{x}} = \frac{1}{\sqrt{1-\mathbf{x}\cdot\mathbf{x}/c^2}}.$$

It is known that the space $\mathbb{R}_c^3$ is closed under the relativistic velocity composition: if $\mathbf{x},\mathbf{y}\in\mathbb{R}_c^3$ then $\mathbf{x}*\mathbf{y}\in\mathbb{R}_c^3$.

For given $\mathbf{a},\mathbf{b}\in\mathbb{R}_c^3$ solve each of the two velocity composition equations

$$\mathbf{a}*\mathbf{x} = \mathbf{b} \tag{1}$$

and

$$\mathbf{x}*\mathbf{a} = \mathbf{b} \tag{2}$$

for the unknown $\mathbf{x}\in\mathbb{R}_c^3$.

6660. *Proposed by Richard Alan Gillman, Valparaiso University, Valparaiso, IN.*

Let f be the multiplicative arithmetic function defined by $f(1) = 1$ and $f(p^a) = pf(a)$ for all primes p and all positive integers a.

(i) Prove that $f(n) \leqslant n$ for all n.

(ii) Prove that the ratio

$$\sum_{j=1}^{n} f(j) \Big/ \sum_{j=1}^{n} j$$

has a limit C as $n \to \infty$.

(iii) Estimate the difference

$$\sum_{j=1}^{n} f(j) - C\sum_{j=1}^{n} j.$$

SOLUTIONS OF ADVANCED PROBLEMS

How Often Is $n\phi(n) \leq x^2$?

6588 [1988, 963]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

As in E3246 put $\psi(n) = n\phi(n)$, where ϕ denotes the arithmetic function of Euler. For positive real x let $F(x)$ denote the number of positive integers n for which $\psi(n) \leq x^2$. Prove that $\lim_{x\to\infty} F(x)/x$ exists and find its value.

Composite Solution by Joachim Herzog, Johann Wolfgang Goethe University, Frankfurt, Germany, Kee-Wai Lau, Hong Kong, and the editors. For $\operatorname{Re} s > 1$ we

velocity composition law is given by the equation

$$\mathbf{x}*\mathbf{y} = \frac{\mathbf{x}+\mathbf{y}}{1+\mathbf{x}\cdot\mathbf{y}/c^2} + \frac{1}{c^2}\cdot\frac{\gamma_{\mathbf{x}}}{\gamma_{\mathbf{x}}+1}\cdot\frac{\mathbf{x}\times(\mathbf{x}\times\mathbf{y})}{1+\mathbf{x}\cdot\mathbf{y}/c^2}, \qquad \mathbf{x},\mathbf{y}\in\mathbb{R}_c^3,$$

where $\gamma_{\mathbf{x}}$ is the *Lorentz factor*

$$\gamma_{\mathbf{x}} = \frac{1}{\sqrt{1-\mathbf{x}\cdot\mathbf{x}/c^2}}.$$

It is known that the space $\mathbb{R}_c^3$ is closed under the relativistic velocity composition: if $\mathbf{x},\mathbf{y}\in\mathbb{R}_c^3$ then $\mathbf{x}*\mathbf{y}\in\mathbb{R}_c^3$.

For given $\mathbf{a},\mathbf{b}\in\mathbb{R}_c^3$ solve each of the two velocity composition equations

$$\mathbf{a}*\mathbf{x} = \mathbf{b} \tag{1}$$

and

$$\mathbf{x}*\mathbf{a} = \mathbf{b} \tag{2}$$

for the unknown $\mathbf{x}\in\mathbb{R}_c^3$.

6660. *Proposed by Richard Alan Gillman, Valparaiso University, Valparaiso, IN.*

Let f be the multiplicative arithmetic function defined by $f(1) = 1$ and $f(p^a) = pf(a)$ for all primes p and all positive integers a.

(i) Prove that $f(n) \leqslant n$ for all n.

(ii) Prove that the ratio

$$\sum_{j=1}^{n} f(j) \Big/ \sum_{j=1}^{n} j$$

has a limit C as $n \to \infty$.

(iii) Estimate the difference

$$\sum_{j=1}^{n} f(j) - C\sum_{j=1}^{n} j.$$

SOLUTIONS OF ADVANCED PROBLEMS

How Often Is $n\phi(n) \leq x^2$?

6588 [1988, 963]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

As in E3246 put $\psi(n) = n\phi(n)$, where ϕ denotes the arithmetic function of Euler. For positive real x let $F(x)$ denote the number of positive integers n for which $\psi(n) \leq x^2$. Prove that $\lim_{x\to\infty} F(x)/x$ exists and find its value.

Composite Solution by Joachim Herzog, Johann Wolfgang Goethe University, Frankfurt, Germany, Kee-Wai Lau, Hong Kong, and the editors. For $\operatorname{Re} s > 1$ we

The result of this problem may also be obtained by the methods of [4], [5], or [6]. An estimation of the difference $F(x) - f(1)x$ can be obtained by the methods of [1], [2], or [3].

REFERENCES

1. R. Balasubramanian and K. Ramachandra, On the numbers of integers n such that $nd(n) \leq x$, *Acta Arithmetica* 49 (1988) 313–322.
2. Michel Balazard and Abdelhakim Smati, Elementary proof of a theorem of Bateman, *Analytic Number Theory, Proceedings of a Conference in Honor of Paul T. Bateman, Progress in Mathematics*, Volume 85, Birkhäuser, 1990, pp. 41–46.
3. Paul T. Bateman, The distribution of values of the Euler function, *Acta Arithmetica*, 21 (1972) 329–345.
4. Robert E. Dressler, A density which counts multiplicity, *Pacific J. Math.*, 34 (1970) 371–378.
5. Paul Erdös, Some remarks on Euler's ϕ-function and some related problems, *Bull. Amer. Math. Soc.* 51 (1945) 540–544.
6. Jeań-Louis Nicolas, Distribution des valeurs de la fonction d'Euler, *L'Enseignement Math.*, 30 (1984) 331–338.

No other solutions were received.

Asymptotic Formula for the Solution of a Difference Equation

6610 [1989, 744]. *Proposed by A. W. Goodman, University of South Florida, Tampa.*

Suppose the sequence $x_1, x_2 \ldots$ is defined by $x_1 = 1$ and $x_n = x_{n-1} + kx_{n-1}^{-c}$ for $n > 1$, where k and c are given positive constants. Find an asymptotic formula for x_n.

Solution (jointly) by H. Köditz, L. Mattner, E. Mues, and M. Reinders, Universität Hannover, Germany. For $n \to \infty$ we have

$$x_n = ((c+1)kn)^{1/(c+1)} + O(1).$$

This is a consequence of the following result.

Let the sequence $\{x_n\}$ be defined by

$$x_1 = 1, \qquad x_{n+1} = x_n + g(x_n), \qquad n \geq 1,$$

where $g(x)$ is a continuous, positive, and nonincreasing function for $x \geq 1$. Let $x(t)$ be the solution of the initial value problem

$$x(1) = 1, \qquad x'(t) = g(x(t)), \qquad t \geq 1.$$

Then $x_n \to \infty$ and $x_n = x(n) + O(1)$ for $n \to \infty$.

To prove this, first observe that with the conditions imposed on g the differential equation has a unique solution $x(t)$ that is defined for all $t \geq 1$. The function $x(t)$ is increasing, and clearly $x(t) \to \infty$ as $t \to \infty$.

For $t \geq 1$ define $p(t)$ by

$$p(t) = x_n + (t-n)(x_{n+1} - x_n), \qquad n \leq t \leq n+1.$$

This is the usual Euler-Cauchy polygonal approximation, with stepsize 1, for the differential equation. Clearly $p(1) = 1$ and

$$p'_-(t) = x_{n+1} - x_n = g(x_n) \geq g(p(t))$$

The result of this problem may also be obtained by the methods of [4], [5], or [6]. An estimation of the difference $F(x) - f(1)x$ can be obtained by the methods of [1], [2], or [3].

REFERENCES

1. R. Balasubramanian and K. Ramachandra, On the numbers of integers n such that $nd(n) \leq x$, *Acta Arithmetica* 49 (1988) 313–322.
2. Michel Balazard and Abdelhakim Smati, Elementary proof of a theorem of Bateman, *Analytic Number Theory, Proceedings of a Conference in Honor of Paul T. Bateman, Progress in Mathematics*, Volume 85, Birkhäuser, 1990, pp. 41–46.
3. Paul T. Bateman, The distribution of values of the Euler function, *Acta Arithmetica*, 21 (1972) 329–345.
4. Robert E. Dressler, A density which counts multiplicity, *Pacific J. Math.*, 34 (1970) 371–378.
5. Paul Erdös, Some remarks on Euler's ϕ-function and some related problems, *Bull. Amer. Math. Soc.* 51 (1945) 540–544.
6. Jeań-Louis Nicolas, Distribution des valeurs de la fonction d'Euler, *L'Enseignement Math.*, 30 (1984) 331–338.

No other solutions were received.

Asymptotic Formula for the Solution of a Difference Equation

6610 [1989, 744]. *Proposed by A. W. Goodman, University of South Florida, Tampa.*

Suppose the sequence $x_1, x_2 \ldots$ is defined by $x_1 = 1$ and $x_n = x_{n-1} + kx_{n-1}^{-c}$ for $n > 1$, where k and c are given positive constants. Find an asymptotic formula for x_n.

Solution (jointly) by H. Köditz, L. Mattner, E. Mues, and M. Reinders, Universität Hannover, Germany. For $n \to \infty$ we have

$$x_n = ((c+1)kn)^{1/(c+1)} + O(1).$$

This is a consequence of the following result.

Let the sequence $\{x_n\}$ be defined by

$$x_1 = 1, \qquad x_{n+1} = x_n + g(x_n), \qquad n \geq 1,$$

where $g(x)$ is a continuous, positive, and nonincreasing function for $x \geq 1$. Let $x(t)$ be the solution of the initial value problem

$$x(1) = 1, \qquad x'(t) = g(x(t)), \qquad t \geq 1.$$

Then $x_n \to \infty$ and $x_n = x(n) + O(1)$ for $n \to \infty$.

To prove this, first observe that with the conditions imposed on g the differential equation has a unique solution $x(t)$ that is defined for all $t \geq 1$. The function $x(t)$ is increasing, and clearly $x(t) \to \infty$ as $t \to \infty$.

For $t \geq 1$ define $p(t)$ by

$$p(t) = x_n + (t-n)(x_{n+1} - x_n), \qquad n \leq t \leq n+1.$$

This is the usual Euler-Cauchy polygonal approximation, with stepsize 1, for the differential equation. Clearly $p(1) = 1$ and

$$p'_-(t) = x_{n+1} - x_n = g(x_n) \geq g(p(t))$$

1974, 187]. It was proposed by J. J. A. M. Brands; the solution published was by F. W. Steutel.

Solved also by Robert A. Agnew, Pull Cull, Thomas N. Delmer, Mordecai E. Falkowitz (Israel), Zachary Franco, Serge Gaignoux (France), Todor G. Genchev (Bulgaria), Murray Klamkin (Canada), Jean-Charles Leccia (France), Jean-Marie Monier (France), Bernard Pinchuk, Arthur J. Rosenthal, Richard Stong, Nora S. Thornber, and Douglas B. Tyler.

Contracting Mappings in a Compact Metric Space

6611 [1989, 744]. *Proposed by O. Hajek and D. Singer, Case Western Reserve University, Cleveland, OH.*

Suppose X is a compact metric space with metric d and suppose $T: X \to X$ is a homeomorphism of X. If $d(x, y) \geq d(Tx, Ty)$ for all x, y in X, prove that T is an isometry (i.e, $d(x, y) = d(Tx, Ty)$ for all x, y in X).

Solution by John Alexopoulos, Kent State University, Ohio. We replace the homeomorphism assumption by the weaker assumption that T is onto. For any fixed x and y in X and any positive integer n there are points x_n, y_n in X such that

$$T^n(x_n) = x, \qquad T^n(y_n) = y.$$

Since X is compact there is a strictly increasing sequence of positive integers $n(k)$ and a pair of points x' and y' in X such that

$$\lim_k x_{n(k)} = x', \qquad \lim_k y_{n(k)} = y'.$$

Now

$$d\big(T^{n(k)}(x'), x\big) = d\big(T^{n(k)}(x'), T^{n(k)}(x_{n(k)})\big) \leq d\big(x', x_{n(k)}\big).$$

Thus $\lim T^{n(k)}(x') = x$ and (similarly) $\lim T^{n(k)}(y') = y$.

Since the sequence $d(T^n(x'), T^n(y'))$ is bounded and decreasing, it converges to $d(x, y)$. Hence

$$\begin{aligned} d(x, y) &= \lim d\big(T^{n(k)}(x'), T^{n(k)}(y')\big) \\ &= \lim d\big(TT^{n(k)}(x'), TT^{n(k)}(y')\big) \\ &= d\big(T(x), T(y)\big). \end{aligned}$$

Editorial comment. Many solvers gave references for this result in the literature, including exercises in the textbooks of Bourbaki, Dugundji, Engelking, and Kaplansky. The two oldest references were to A. Lindenbaum, Contribution à l'étude de l'éspace metrique, *Fund. Math.*, 8 (1926) 209–222, and H. Freudenthal and W. Hurewicz, Dehnungen, Verkürzungen, Isometrien, *Fund. Math*, 26 (1936) 120–122. A. Witkowski (Poland) mentioned that the result is the case $G = T^{-1}$ of the following. If X is a compact metric space, $G: X \to X$ is continuous, and

$$d(Gx, Gy) \geq d(x, y)$$

for all x and y in X, then G is an isometry. He informs us that the folklore version of the above is "if a hippopotamus closes its eye, its skin must expand somewhere."

Solved also by the proposer and 28 others.

1974, 187]. It was proposed by J. J. A. M. Brands; the solution published was by F. W. Steutel.

Solved also by Robert A. Agnew, Pull Cull, Thomas N. Delmer, Mordecai E. Falkowitz (Israel), Zachary Franco, Serge Gaignoux (France), Todor G. Genchev (Bulgaria), Murray Klamkin (Canada), Jean-Charles Leccia (France), Jean-Marie Monier (France), Bernard Pinchuk, Arthur J. Rosenthal, Richard Stong, Nora S. Thornber, and Douglas B. Tyler.

Contracting Mappings in a Compact Metric Space

6611 [1989, 744]. *Proposed by O. Hajek and D. Singer, Case Western Reserve University, Cleveland, OH.*

Suppose X is a compact metric space with metric d and suppose T: $X \to X$ is a homeomorphism of X. If $d(x, y) \geq d(Tx, Ty)$ for all x, y in X, prove that T is an isometry (i.e, $d(x, y) = d(Tx, Ty)$ for all x, y in X).

Solution by John Alexopoulos, Kent State University, Ohio. We replace the homeomorphism assumption by the weaker assumption that T is onto. For any fixed x and y in X and any positive integer n there are points x_n, y_n in X such that

$$T^n(x_n) = x, \qquad T^n(y_n) = y.$$

Since X is compact there is a strictly increasing sequence of positive integers $n(k)$ and a pair of points x' and y' in X such that

$$\lim_k x_{n(k)} = x', \qquad \lim_k y_{n(k)} = y'.$$

Now

$$d\big(T^{n(k)}(x'), x\big) = d\big(T^{n(k)}(x'), T^{n(k)}(x_{n(k)})\big) \leq d\big(x', x_{n(k)}\big).$$

Thus $\lim T^{n(k)}(x') = x$ and (similarly) $\lim T^{n(k)}(y') = y$.

Since the sequence $d(T^n(x'), T^n(y'))$ is bounded and decreasing, it converges to $d(x, y)$. Hence

$$\begin{aligned} d(x, y) &= \lim d\big(T^{n(k)}(x'), T^{n(k)}(y')\big) \\ &= \lim d\big(TT^{n(k)}(x'), TT^{n(k)}(y')\big) \\ &= d\big(T(x), T(y)\big). \end{aligned}$$

Editorial comment. Many solvers gave references for this result in the literature, including exercises in the textbooks of Bourbaki, Dugundji, Engelking, and Kaplansky. The two oldest references were to A. Lindenbaum, Contribution à l'étude de l'éspace metrique, *Fund. Math.*, 8 (1926) 209–222, and H. Freudenthal and W. Hurewicz, Dehnungen, Verkürzungen, Isometrien, *Fund. Math*, 26 (1936) 120–122. A. Witkowski (Poland) mentioned that the result is the case $G = T^{-1}$ of the following. If X is a compact metric space, G: $X \to X$ is continuous, and

$$d(Gx, Gy) \geq d(x, y)$$

for all x and y in X, then G is an isometry. He informs us that the folklore version of the above is "if a hippopotamus closes its eye, its skin must expand somewhere."

Solved also by the proposer and 28 others.

Mahler's Inequality

6613 [1989, 846]. *Proposed by Jeff Vaaler, University of Texas, Austin.*

Suppose P is a polynomial of degree n with complex coefficients, leading coefficient a and zeros $\alpha_1, \ldots, \alpha_n$. Define the Mahler measure $M(P)$ of P by

$$M(P) = \exp \int_0^1 \log|P(e^{2\pi it})|\, dt = a \prod_{j=1}^{n} \max(1, |\alpha_j|),$$

where the latter equality follows by Jensen's formula. If P' is the derivative of P, prove that

$$M(P') \leq nM(P).$$

Solution by David W. Boyd, University of British Columbia, Vancouver. First observe that if $Q(z)$ is a polynomial of degree n with leading coefficient b having all its zeros inside the unit circle $|z| \leq 1$, then $M(Q) = |b|$. The leading coefficient of Q' is then nb and all zeros of Q' lie inside the unit circle, by the Gauss-Lucas theorem (see p. 22 of Morris Marden, *Geometry of Polynomials*, Mathematical Surveys No. 3, Amer. Math. Soc., 1966). Thus

$$M(Q') = n|b| = nM(Q),$$

so the desired inequality holds with equality for such Q.

Now suppose P is an arbitrary polynomial of degree n. Then there is a polynomial Q of degree n with all zeros in $|z| \leq 1$ and such that $|Q(z)| = |P(z)|$ for all $|z| = 1$. Simply define

$$Q(z) = P(z) \prod_{|\alpha_j|>1} \frac{1 - \bar{\alpha}_j z}{z - \alpha_j}.$$

Then Q has the property claimed since $|(1 - \bar{\alpha}_j z)/(z - \alpha_j)| = 1$ for all $|z| = 1$. The integral formula for the measure then shows that $M(Q) = M(P)$.

Finally, we recall an important theorem of Bernstein (ibid, p. 23) which states that if $|P(z)| \leq |Q(z)|$ for all $|z| = 1$, and if $Q(z)$ does not vanish outside the unit circle, then $|P'(z)| \leq |Q'(z)|$ for all $|z| = 1$. Thus

$$M(P') \leq M(Q') = nM(Q) = nM(P).$$

Remark 1. Mahler (*Proc. Roy. Soc. Lond. Ser. A*, 264 (1961) 145–154) gave a rather involved proof of this inequality. In a postscript, he pointed out that a referee had remarked that the result could be proved by observing that

$$\log \frac{M(P')}{M(P)} = \int_0^1 \log \left| \sum_{j=1}^{n} \frac{1}{e^{2\pi it} - \alpha_j} \right| dt$$

is a subharmonic function of the zeros of P and hence the maximum is attained when they lie on the unit circle. The proof of the theorem of Bernstein used in the above solution uses only the maximum modulus principle for the function $P(z)/Q(z)$ and the Gauss-Lucas theorem, and hence the above proof is somewhat more elementary.

Remark 2. For $0 < p < \infty$, let

$$\|P\|_p = \left\{ \int_0^1 |P(e^{2\pi it})|^p \, dt \right\}^{1/p}$$

be the L^p mean of P on the unit circle. Then $\|P\|_p \to M(P) = \|P\|_0$ as $p \to 0$ and

$$\|P\|_p \to \max_{|z|=1} |P(z)| = \|P\|_\infty$$

as $p \to \infty$. The well known inequality of Bernstein states that $\|P'\|_\infty \leq n\|P\|_\infty$ if P is a polynomial of degree n. This was extended to $\|P'\|_p \leq n\|P\|_p$ for $p \geq 1$ by Zygmund (*Trigonometric Series*, Cambridge University Press, Volume II, 1968 p. 11). The inequality of Mahler, proved above, shows that this holds also for $p = 0$.

It is conjectured that the inequality $\|P'\|_p \leq n\|P\|_p$ continues to hold for $0 < p < 1$. In this range, the best result known is due to Máté and Nevai (*Annals of Math.*, 111 (1980) 145–154), who showed that the inequality holds with the factor $n(4e)^{1/p}$ in place of n.

One incomplete solution was also received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Representations and Characters of Finite Groups. By M. J. Collins, Cambridge University Press, 1990. xii + 242 pp.

MARK STEVEN MAZUR
Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282

Representation theory and character theory have been a major tool in the study of finite groups. Without them the classification of the finite simple groups might never have been achieved, or at the very least, might still be underway. The author of this text is keenly aware of the central role that representation and character theory have played in the classification problem. In fact, the main objective of his presentation is a development of the representation and character theory necessary for tackling the classification problem.

Representation and character theory began with Frobenius's work, nearly a century ago. His theorem on transitive permutation groups along with Burnside's theorem on the solvability of groups of order $p^a q^b$ are major results in the theory. While more recently a proof of Burnside's theorem has been given which is purely group theoretic, no such proof of Frobenius's result exists. Notable, too, is Feit and Thompson's odd order paper (the solvability of groups of odd order) which relies upon the developments of representation and character theory. Unlike many mathematicians, the author of this book has gone to considerable lengths to inform the reader of the theory's historical importance and of the scholars who have figured significantly in the development of the theory; he ascribes the important theorems to their authors, as is customary, but also includes the year in which results were published. Such a technique gives even the casual reader a clear picture of how the many threads of the theory developed and how they were woven together and applied to the classification problem.

Stated as simply as possible, the basic objects of study are representations and characters. A representation is a homomorphism from an abstract group G to the general linear group $GL(V)$ where V is a finite-dimensional vector space over a field K. If the dimension of V is n and a basis is chosen for V, then V may be regarded as the vector space of n-tuples over K, and $GL(V)$ may be viewed as the group $GL(n, K)$ of nonsingular $n \times n$ matrices over K. Historically, these matrix representations of a group were considered first since the groups themselves were well understood. For each representation of a group there is an associated character defined as the trace of the matrix representation. These characters are class functions, because they are constant on the conjugacy classes of the group. Hence, the study of the characters of a finite group is often sufficient even though information is sacrificed (since only the trace of the matrix is used rather than the entire matrix itself).

In terms of the classification problem one often wishes to deduce global information about the group G from local information (i.e., from knowledge of certain subgroups H of G). Many results in representation and character theory

give conditions under which representations and characters of a subgroup may be "lifted" to the group. This sometimes forces the group to have a particular isomorphism type. Such a strategy is employed often in the classification problem.

Chapter 1 of Collins's book contains an exposition of representation theory. Chapter 2 develops character theory. These two chapters contain all the standard basic theory from these areas. The author also includes a few specialized results which he will have occasion to apply later in the book.

The remainder of the book is more specialized. Chapter 3 contains Suzuki's theory of exceptional characters. This chapter deals with conditions under which the knowledge of the characters of a subgroup H of G provides information about the characters of a group G. As mentioned before, this knowledge may force G to have a certain isomorphism type. Chapter 4 discusses coherence and the exceptional characters. The chapter's goal is to study CN-groups of odd order and Zassenhaus groups. The study of these groups was important in the early stages of the classification of the finite simple groups. Chapter 5 contains Brauer's characterization of characters which gives necessary and sufficient conditions under which a complex-valued class function is a character in terms of the elementary subgroups of the group. Collins follows the Brauer-Tate proof rather than the more recent Goldschmidt-Isaacs proof. Leaving the characterization of characters to this late stage in a book on character theory is unusual. The author delays its inclusion because he makes a very interesting use of this result in chapter 6. He introduces block theory in a nonstandard way. For instance, he obtains a proof of the nonsimplicity of groups with homocyclic Sylow 2-subgroups of rank 2 by means of an isometry constructed using Brauer's characterization of characters rather than using Brauer's method of columns. Because of the success in applying character ring methods to situations in which modular characters have been applied in the past, the author speculates in the book's preface that perhaps one day non-modular proofs may be obtained for Brauer's second and third main theorems which are fundamental results in the area of modular characters.

The author concludes the book with a brief appendix containing Glauberman's Z^*-theorem because of its importance to the classification problem.

In summary, Collins's book is a well-written text with a very clear purpose in mind—a self-contained development of representation and character theory with the classification of finite simple groups as the driving force behind the development. The book is written for the specialist or the graduate student with a strong background in algebra. I must disagree with the author who says that the text could be used with gifted undergraduates if "a judicious choice of material" is made. I certainly do not believe that most American undergraduates would possess the necessary background.

The value of this book lies not in its encyclopedic approach to the subject—Curtis and Reiner is the standard here—nor in its plethora of examples —the text is written concisely. The text's greatest attraction lies in its novel approach to the theory in light of the classification problem.

Theorist. By Prescience Corporation, 814 Castro Street, San Francisco, CA 94114.

FRANK WATTENBERG
Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003

The Prescience Corporation describes *Theorist* as "the first symbolic algebra system to make computer mathematics easy." The following table was included with *Theorist*.

program name	retail price	recommended memory	minimum memory	required disk space
Theorist	$379	1000k	600k	700k
Mathematica	$795*	4000k	2500k	4800k
Maple	$395	2000k	1000k	7000k
Milo	$249	512k	384k	400k

*Macintosh II version sold separately from regular version, which is $495.
Other programs' prices include both versions.

With the list price of the new Mac Classic under $1,000 is *Theorist* the program we've all been waiting for? Do we and our students now have affordable access to a serious computer-based mathematical environment with graphics, numerics and symbolic algebra? Unfortunately, the answer is no. Although *Theorist* does some things very well, it has shortcomings that severely limit its usefulness to both working mathematicians and students.

This review compares *Theorist* to Mathematica and the Hewlett-Packard HP-48SX calculator. Except for the unavoidable limitations of its checkbook size—its display is only 131 pixels by 64 pixels, and its tiny keyboard—the HP-48SX is a worthy, affordable, and portable competitor. In this league, in addition to the usual desiderata for computer programs—power, flexibility, reliability, robustness, and user-friendliness—we look for

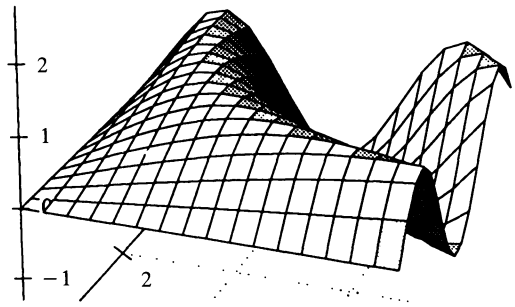
- An integrated environment for doing mathematics.
- The ability to treat problems in a natural and direct way.
- The ability to construct powerful mathematical modules.

Running *Theorist* after using Mathematica is like driving a bright red, two-seater sports car—quick, lively, and eye-catching. From a standing start on the hard drive it's up and ready to go in six seconds[1]. The much larger Mathematica requires more than four times as long. *Theorist* graphs the function $y = \sin 5t/2^t$ in under two seconds; Mathematica requires eight seconds; and the HP-48SX requires 22 seconds for the same graph.

Theorist is optimized for speed and efficient use of resources. For example, its default values produce graphs with relatively few subdivisions and its goal for

[1]Timings for *Theorist* and Mathematica are based on a Mac IICX.

calculations is 19 digits of precision,[2] sufficient for all but the most unusual circumstances and faster than the "arbitrary precision" claimed by other programs. *Theorist* produced the figure below in just a few seconds.



$z = \sin xy$

The speed and responsiveness of *Theorist*'s graphics make for a wonderful interactive graphing environment in which it is easy and enjoyable to adjust viewing angles and perspective in real time to produce good-looking and informative 3D graphs. Mathematica's graphics are more powerful but its user interface is not as responsive as *Theorist*'s. For example, in *Theorist* one adjusts the viewing angle for a 3D graph by clicking and dragging a box enclosing the graph. Both programs have the ability to produce animated sequences, although Mathematica's is much more flexible. The HP-48SX, of course, is not in the same graphics league. Its size rules out a truly useful graphics screen, but it can produce rough 2-dimensional graphics suitable for experimentation and has surprisingly powerful and useful graphics commands.

Formulas in *Theorist* can be entered with enough pointing-and-clicking to keep the most hyperactive mouse happy although the documentation wisely encourages users to switch quickly to the faster keyboard entry. While Mathematica's formula display harkens back to the days of teletypes, both *Theorist* and the HP-48SX display formulas the way we usually write them.

Mathematica and the HP-48SX have extensive and well thought-out programming capabilities. Both encourage top-down programming and have intelligent and powerful data types. Both can treat functions as easily as other data types and both have much of the power and elegance of LISP. Using either the HP-48SX or Mathematica one rapidly builds up a library of powerful programs and, hence, of powerful concepts that enable one to deal with mathematics on a new level. This is where *Theorist* falls short. It has only a few very limited programming capabilities and one is forced to deal with mathematics at a very basic level.

The HP-48SX and Mathematica encourage the treatment of problems in a natural and clean way. Learning either one is mentally stimulating and intellectually rewarding. Both provide powerful and often new ways of conceptualizing

[2]But on page 19 of the *Theorist* reference manual we find "To avoid unreasonable performance, we had to limit the precision of some algorithms to much less than nineteen digits, although the result should be sufficient for most purposes."

mathematical problems. By contrast *Theorist* is stuck on a high school view of mathematics as mechanical symbolic manipulation. For example, the *Theorist* learning guide devotes eleven pages to integrating

$$\int \sqrt{a + b \sin^2 \theta}\, \cos \theta \, d\theta$$

with the user guiding *Theorist* step-by-step through all the gory details. With Mathematica one types

```
Integrate[Sqrt[a + b Sin[t]^2] Cos[t], t]
```

and in less than three seconds

$$\frac{a \log\left(\sqrt{b}\, \sin(t) + \sqrt{a + b \sin(t)^2}\right)}{2\sqrt{b}} + \frac{\sin(t)\sqrt{a + b \sin(t)^2}}{2}$$

appears.

Case Studies: Doing Mathematics

The Fibonacci sequence. The Fibonacci sequence is one of the standard examples illustrating the power and computational pitfalls of recursion. It is discussed in the documentation for all three of the systems under discussion. The sequence is defined by

$$F_0 = F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}.$$

The obvious program in any language

```
Fib[0] = 1
Fib[1] = 1
Fib[n_] := Fib[n - 1] + Fib[n - 2]
```

runs in exponential time since it requires the computation of F_{n-1} and of F_{n-2} from scratch to determine F_n. This problem is very general. For example, the obvious program for simulating the discrete logistic model

$$p_n = ap_{n-1}(1 - bp_{n-1})$$

also requires exponential time since it makes two function calls for p_{n-1} in the computation of p_n. With both Mathematica and the HP-48SX there are easy ways to avoid this and these ways are discussed in the documentation. In Mathematica, for example, the construction

```
Fib[n_] := Fib[n] = Fib[n - 1] + Fib[n - 2]
```

saves the result of each computation for future use. Perhaps the best way to look at the problem for the logistic model is by rewriting it in the form

$$f(p) = ap(1 - bp)$$

$$p_n = f(p_{n-1})$$

The obvious program corresponding to this formulation

```
F[p_] := a p (1 - b p)
P[n_] := F[P[n - 1]]
```

runs very quickly and focuses attention where it belongs—on iteration of the function $f(p)$.

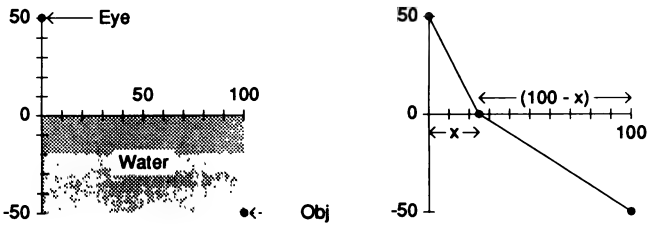
We can treat the Fibonacci sequence in the same way using vectors.

$$\vec{G}(\vec{x}) = (x_2, x_1 + x_2)$$
$$\vec{v}_2 = (1, 1)$$
$$\vec{v}_n = \vec{G}(\vec{v}_{n-1}).$$

Besides solving the computation problem this solution focuses attention on the two-dimensional nature of the problem. This solution is very natural using either Mathematica or the HP-48SX and appears to be the only solution using *Theorist*[3], but the *Theorist* reference manual gives up. It suggests computing the Fibonacci sequence by using the obvious recursion to compute F_8 and F_9 and then adding these as "initial conditions" to compute the next few terms in the sequence. In this way one could theoretically compute as many terms as desired, seven at a time. The following *Theorist* notebook shows the more elegant solution.

$$f(q) = \begin{pmatrix} q_2 \\ q_1 + q_2 \end{pmatrix}$$

$$\text{fib}(n) = \begin{cases} \begin{pmatrix} 1 \\ 1 \end{pmatrix} & (n \le 2) \\ f(\text{fib}(n-1)) & (2 < n) \end{cases}$$



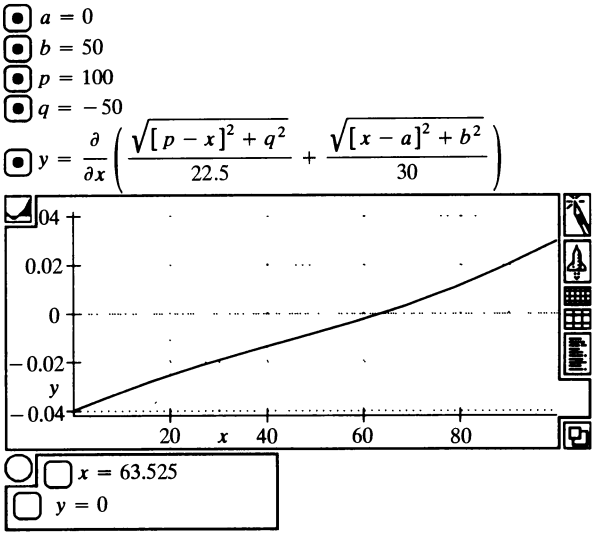
Refraction problem

A Typical Calculus Problem—Refraction. Suppose that we want to calculate the path followed by a light ray traveling from an underwater object to an abovewater eye as shown in the figure above. More generally, we can suppose that the eye is located at (a, b) and the object at (p, q). According to Fermat's Principle, light will follow the path that minimizes the total travel time from the object to the eye. Since the speed of light in air is roughly 30 cm/ns and the speed of light in water is roughly 22.5 cm/ns we find this point by minimizing the function

$$T(x) = \frac{\sqrt{(p-x)^2 + q^2}}{22.5} + \frac{\sqrt{(x-a)^2 + b^2}}{30}.$$

[3]*Theorist*, however, has a very low limit for the depth of recursion.

The following is a *Theorist* notebook solving this problem.



This notebook does not show the required user interaction. The user must request a graph by pointing-and-clicking, then choose suitable x and y limits (*Theorist* does not even provide automatic y-axis scaling) and, having visually isolated the root, request a root. The following printout shows a Mathematica notebook solving the same problem.

```
T[x_, a_, b_, p_, q_] := Sqrt[(p - x) ^2 + q ^2] / 22.5 +
                         Sqrt[(x - a) ^2 + b ^2] / 30
Tprime[x_, a_, b_, p_, q_]   := D[T[x,a,b,p,q], x]
Answer[a_, b_, p_, q_]:=
    FindRoot [Tprime[x, a,b,p,q] == 0, {x, (a + b) / 2}]

Answer[0, 50, 100, - 50]
{x -> 63.5253}
```

These two solutions show some of the differences between Mathematica and *Theorist*. The Mathematica solution is broken down in the natural way

- Define the function $T(x)$ that describes the total travel time if the light ray intersects the surface of the water at the point x.
- Differentiate the function $T(x)$.
- Solve the equation $T'(x) = 0$.

The structure of Mathematica encourages this solution because of the facility with which it handles functions. *Theorist* handles functions in a very muddy way. One cannot in a natural way define a function $T(x)$, then differentiate it to obtain $T'(x)$, and then solve the equation $T'(x) = 0$ as we did in Mathematica. Indeed, *Theorist* cannot differentiate functions—only expressions. It requires that we define a new variable y in one line by differentiating the expression defining the function $T(x)$.

All the user involvement with *Theorist* is nice the first few times but it prevents progressing further. For example, I like to ask my students to solve real problems. One particularly nice problem is finding the apparent position and shape of an underwater object as viewed by an abovewater person using binocular vision[4]. With Mathematica or the HP-48SX it is easy and natural to solve more difficult problems like these using the function

```
answer[a,b,p,q]
```

as a building block. With *Theorist* a student would spend most of his or her time repeating the same sequence of calculations.

Conclusions

Although *Theorist* is a worthwhile program with fast and powerful graphics which are a joy to use interactively, both the HP-48SX and Mathematica are more powerful environments for doing useful mathematics. Mathematics progresses in large part by building new concepts that allow us to work in broader strokes. Because it is impossible to build mathematical modules within *Theorist*, the user is confined to a basic level. *Theorist* is not flexible, in large part because it lacks powerful and general programming constructs. It is often difficult or impossible to approach a problem in *Theorist* in the most natural way. *Theorist* does symbolic manipulation well but because of the massive user interaction required to solve even the most straightforward problems, this ability is useful for solving schoolbook problems rather than doing serious work.

For fast, responsive graphics and interactive mechanical algebraic manipulation choose *Theorist*. For serious work choose Mathematica or the HP-48SX.

[4]If the viewer's two eyes and the object determine a plane perpendicular to the surface of the water then the apparent position of the object is shifted toward the viewer and upward. If the viewer's two eyes are the same height above the surface of the water and the same distance from the object then the apparent position of the object is shifted directly upward from its actual position. Generically, the two apparent lines of sight do not intersect. The mathematics predicts interesting optical phenomena easily observable with a ten gallon aquarium.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, T(14), S, C, L. *Microcomputers and Mathematics.* J.W. Bruce, P.J. Giblin, P.J. Rippon. Cambridge Univ Pr, 1990, xvi + 425 pp, $90; $29.95 (P). [ISBN: 0-521-37515-0; 0-521-31238-8] An interesting collection of topics to be investigated by students using computers. The mathematical prerequisites vary from high school algebra up through calculus, and is accessible to freshmen and sophomores. Topics include elementary number theory, curves, differential equations, and iteration of functions. Abundant examples, exercises, and sample computer programs. The programs are written in Basic, but are easily transferrable to other environments (such as *Mathematica*). Overall, an excellent reference book for anyone wishing to incorporate computing into the lower-level mathematics curriculum. Note the price. MPR

General, P, L.** *A.J. Lohwater's Russian-English Dictionary of the Mathematical Sciences, Second Edition, Revised and Expanded.* Ed: R.P. Boas. AMS, 1990, xi + 343 pp, $35 (P). [ISBN: 0-8218-0133-3] Long-awaited update of this important reference work, first published thirty years ago. Opening grammar synopsis has been rewritten; stress marks have been added to Russian words; and the vocabulary has been "extensively enlarged" to reflect contemporary Russian writing in the mathematical sciences. Intended only to aid translation from Russian to English, not vice versa. LAS

General, C, P. *User Manual for DERIVE, Version 2: A Mathematical Assistant for Your Personal Computer.* (IBM PC Software.) Albert Rich, Joan Rich, David Stoutemyer. Soft Warehouse, 1990, viii + 244 pp, $250 (P). The most significant enhancement in this new version is programmability. Users can now augment the collection of built-in functions with user-defined functions. The new programming capability allows the definition of iterative and recursive functions. In addition to internal improvements in the software, the *User Manual* has been extensively revised and expanded. Despite the increased functionality, *DERIVE* continues to run on any IBM PC or compatible computer with at least 512K of memory and MS-DOS 2.1 or later. AO

Elementary, T(13: 1). *Elementary Geometry, Third Edition.* R. David Gustafson, Peter D. Frisk. Wiley, 1991, ix + 446 pp, $43.95. [ISBN: 0-471-51002-5] Changes in this edition include revision of the chapter on logic, additional material on analytic geometry, surface area and volume, more algebraic problems, and new end-of-chapter review exercises. Remains a very traditional presentation of high school Euclidean geometry, complete with two-column proofs. (*First Edition,* TR, April 1973; *Second Edition,* TR, February 1986.) JNC

Elementary, S*(7-13). *Polynomials.* Tom M. Apostol. Project Mathematics! California Institute of Technology, 1991, 28 minute videotape. *Program Guide and Workbook,* 30 pp, (P). A half-hour videotape motivated by basketball shots that uses effective video graphics to illustrate linear, quadratic, cubic, and higher polynomials. An animated hand crank simultaneously varies parameters and shapes of graphs; on-screen algebraic choreography helps relate algebraic forms to visual representations, especially of factors and zeros. Accompanying *Workbook* provides reinforcement, extensions, and exercises. Available from state Departments of Education, from the MAA, and from the Cal Tech bookstore; freely reproduceable for educational purposes. LAS

Mathematics Appreciation. *For All Practical Purposes: Introduction to Contemporary Mathematics, Second Edition.* COMAP. WH Freeman, 1991, xvii + 626 pp. [ISBN: 0-7167-2115-5] The new edition of this popular liberal arts survey of contemporary mathematics contains a welcome doubling of the exercise sets, together with key revisions in selected chapters. The optional accompanying videotape series (not revised) remains well-coordinated with the text. Updated instructor, telecourse guides available. (*First Edition,* TR, February 1988; Extended Review, May 1989.) RB

Mathematics Appreciation, S, C. *Cube and Tess.* (Macintosh Software.) Atlantic Software (PO Box 299, Wenham, MA 01984), 1988, 21 pp, $24.95

(P). A computer simulation of Rubik's Cube and a four-dimensional version called Tess (for tesseract). Moves are made by clicking; sequences of moves can be recorded into procedures for repeated execution and saved in files for later use. Includes two such files of pre-written procedures for each puzzle. Runs on any Macintosh. LAS

Mathematics Appreciation, S(9-13). *The Story of π.* Tom M. Apostol. Project Mathematics! California Institute of Technology, 1989, 24 minute videotape. *Program Guide and Workbook,* 30 pp, (P). A visual, historical introduction to π, produced in a thoroughly professional manner. Simple applications, Archimedes' discovery, computation, and extensions (lattices, random numbers, Buffon needle). Useful at many levels, from junior high to college classes in quantitative literacy. *Workbook* contains commentary and exercises. Available from state Departments of Education, from MAA, and from the Cal Tech bookstore. LAS

Mathematics Appreciation, T(13-15: 1). *The Nature of Mathematics, Sixth Edition.* Karl J. Smith. Brooks/Cole, 1991, xv + 816 pp. [ISBN: 0-534-13914-0] A warm, "student-friendly" approach to mathematical literacy (reasoning, computers, numbers, algebra, geometry, probability, statistics) for those who learned little (and remember less) of high school mathematics. Three problem levels make the text suitable for courses at a wide range of prerequisites, from none to intermediate algebra. Countless marginal asides enrich the exposition with historical, social, and humorous tidbits. Numerous supplements are available to aid instructors. LAS

Precalculus, S*, L. *Functions and Graphs.* I.M. Gel'fand, E.G. Glagoleva, E.E. Shnol. Birkhauser Boston, 1990, ix + 105 pp, $14.50 (P). [ISBN: 0-8176-3532-7] First in a series of self-study booklets written in 1966 (in Russian) to support Moscow State University's mathematical "Correspondence School," an outreach effort to students aged 12-17 in small, remote villages of the USSR. This pamphlet is designed to help students learn to "see" formulas and functions and observe the way changes in functions are reflected in changes in their graphs. First translated into English by the University of Chicago; original English edition was published in 1967 by MIT Press (TR, October 1969). This new printing makes these attractive books available to a new generation of students. LAS

Education, P*, L*. *Professional Standards for Teaching Mathematics.* NCTM, 1991, viii + 196 pp, $25 (P). [ISBN: 0-87353-307-0] A companion to NCTM's influential 1989 *Standards* for curriculum and evaluation (TR, June-July 1989), this volume describes 24 standards for teaching mathematics, for evaluation of the teaching of mathematics, and for the support and development of mathematics teachers. Emphasizes the need for teachers—and the colleges who educate teachers—to create classroom communities, to promote reasoning, to encourage conjecturing, and to build connections. Expounds a bold vision of what teachers should know and be able to do, how they should be educated, evaluated, and sustained as professionals. LAS

Education, S(17-18). *Algebra for Everyone.* Ed: Edgar L. Edwards, Jr. NCTM, 1990, vi + 89 pp, $7 (P). [ISBN: 0-87353-297-X] Seven essays based on the premise that algebra can and should be taught to all students. Explores issues, delineates fundamental algebraic concepts and processes, and provides teaching ideas. Includes annotated resource list organized into categories that include expectations, equity, learning styles, cooperative learning, materials, manipulatives, software, successful programs, and funding sources. MW

Education, P, L. *Technology in Education: Looking Toward 2020.* Raymond S. Nickerson, Philip P. Zodhiates. Lawrence Erlbaum Assoc, 1988, xviii + 330 pp, $22.50 (P); $59.95. [ISBN: 0-8058-0297-5; 0-8058-0214-2] A collection of fourteen papers on the long-term role of technology in the teaching of mathematics, science, and computing at the primary and secondary school levels. The papers in this volume were originally prepared for a conference held October 15-17, 1986. AO

Education, T(16-17). *Mathematics for the Young Child.* Ed: Joseph N. Payne. NCTM, 1990, xii + 306 pp, $39.50. [ISBN: 0-87353-288-0] Research-based strategies for teaching mathematics to children from pre-school through grade four. Major goals are to enable children to construct their own knowledge and understandings, and to develop procedural knowledge from a strong conceptual knowledge base. Makes connections with recent research, reform documents. Covers whole number concepts, fractions and decimals, geometry and measurement, probability, graphing, and microcomputers, all from a problem-solving and constructivist perspective. MW

History, S(17-18), P, L. *Ein Jahrhundert Mathematik 1890-1990.* Gerd Fischer, *et al.* Dokumente zur Geschichte der Mathematik, Band 6. Friedr Vieweg & Sohn, 1990, xii + 830 pp, DM 198. [ISBN: 3-528-06326-2] Nineteen papers surveying the development of eighteen special fields and the history of the Deutsche Mathematiker Vereinigung during its first hundred years. Contains a fascinating account of mathematics under the Nazis. JD-B

History, P, L*. *John von Neumann and the Origins of Modern Computing.* William Aspray. Hist. of Comput. MIT Pr, 1990, xvii + 376 pp, $35. [ISBN: 0-262-01121-2] A revealing archive-based portrait of the many dimensions of von Neumann's pioneering work in computing. As a central leader in the complex post-war interactions of science, military, and industry, von Neumann not only laid down the principles for stored program computers, but he also helped initiate theories of automata, reliability, self-replication, and memory in biological systems. Text reveals close ties to numerical meteorology and applied mathematics. Well-illustrated with photographs from family and government archives. LAS

History, P, L*. *An Introduction to the His-*

tory of Structural Mechanics. Edoardo Benvenuto. Springer-Verlag, 1991, $49 each. *Part I: Statics and Resistance of Solids,* xxi + 306 pp, [ISBN: 0-387-96227-1]; *Part II: Vaulted Structures and Elastic Systems,* xxi + 245 pp. [ISBN: 0-387-97187-4] An outgrowth of but not a translation of *La Scienza delle Costruzioni e il suo Sviluppo Storico,* Sansoni Publishers, 1981. In a Foreward, Clifford Truesdell says, "The first to show how statics, strength of materials, and elasticity grew alongside existing architecture with its millenial traditions, its host of successes, its ever-renewing styles, and its numerous problems of maintenance and repair—(Benvenuto) thinks clearly, clearly organizes his material, difficult and complicated as it seems, and writes clearly with direct and masterly expression—a general, pioneering treatise—any place you open this book and read in it, you will be fascinated by what is there." The translation from Italian to English is especially well done, without loss of charm. JDEK

Logic, P, L. *The Semantic Foundations of Logic, Volume 1: Propositional Logics.* Richard L. Epstein. Nijhoff Intern. Philosophy Ser., V. 35. Kluwer Academic, 1990, xxii + 386 pp, $99. [ISBN: 0-7923-0622-8] Presents history, philosophy, and mathematics of major systems of propositional logic: classical logic, dependence logics, modal logic, many-valued logics, intuitionism, paraconsistent logic. Proposes general framework for semantics for propositional logics and theory of translations between logics. KES

Graph Theory, P. *Field Guide to Simple Graphs.* Peter Steinbach. Design Lab (6101 Sequoia, NW, Albuquerque, NM 87120), 1990, vii + 134 pp, $20 (P). Drawings of all nonisomorphic, simple, connected graphs with at most seven vertices and all such regular graphs with at most ten vertices, each with basic properties identified. Drawings of all trees with at most twelve vertices; summary tables of numbers of graphs with specified property. Drawings generated "by eye." JPH

Graph Theory, P, L. *Paths, Flows, and VLSI-Layout.* Eds: Bernhard Korte, *et al.* Algorithms & Combinatorics, V. 9. Springer-Verlag, 1990, xxii + 383 pp, $69. [ISBN: 0-387-52685-4] Survey chapters based on a 1988 workshop on graph-theoretic aspects of very-large-scale-integrated circuits. BC

Combinatorics, P. *Mathematics of Ramsey Theory.* Eds: Jaroslav Nešetřil, Vojtěch Rödl. Algorithms and Combinatorics, V. 5. Springer-Verlag, 1990, xiv + 269 pp, $69. [ISBN: 0-387-18191-1] Expository surveys and research on classical and recent results in Ramsey theory of sets (finite and infinite), graphs, numbers, Euclidean geometry, and topological spaces. Also complexity and independence results, and techniques of ergodic theory. Articles by leaders in the field. JPH

Combinatorics, P. *Finite Geometries and Combinatorial Designs.* Eds: Earl S. Kramer, Spyros S. Magliveras. Contemp. Math., V. 111. AMS, 1990, xviii + 312 pp, $53. [ISBN: 0-8218-5118-7] Proceedings of an AMS special session held in Lincoln, Nebraska during October 29-November 1, 1987. Contains 23 papers. AD

Number Theory, T*(15-18: 1, 2), S*, L.** *An Introduction to the Theory of Numbers, Fifth Edition.* Ivan Niven, Herbert S. Zuckerman, Hugh L. Montgomery. John Wiley, 1991, xiii + 529 pp, $50.95. [ISBN: 0-471-62546-9] A classic re-born. Earlier editions appeared in 1960 (original), 1966 (*Second*), 1972 (*Third,* TR, June-July 1972), and 1980 (*Fourth,* TR, December 1980). This edition reflects many of the remarkable new developments in number theory of the past decade (e.g., public key cryptography, elliptic curves and Falting's theorem, pseudoprime tests, Lenstra's factorization) as well as other enrichments (e.g., geometry of numbers, systems of linear Diophantine equations, numerous calculational issues). Extensive notes tie each section to historical and bibliographic roots. Can be used either for a one- or two-term course: selected early sections provide a gentle pace suitable for novices. LAS

Number Theory, S(15-16), C, P. *Number Theory Package.* Donald E.G. Malm. Oakland University (E-mail: malm@argo.acs.oakland.edu), 1990, (P). Three implementations (for Turbo Pascal on IBM clones and on Macintosh, and for UBASIC on IBM clones) of a suite of forty research-level subroutines and function calls for number theory, implementing many current algorithms for efficient computation. Includes examples showing how the routines can be put together, and a list of all procedures with descriptions and limitations. Samples: Baille-Wagstaff Lucas pseudo-prime test; elliptic curve method to factorize; Lehmer's algorithm for linear congruences. Preferred distribution by anonymous ftp to `vela.acs.oakland.edu`, in /pub/numbdroid subdirectory. Also available on disk. LAS

Linear Algebra, P, L. *Nonnegative Matrices in Dynamic Systems.* Abraham Berman, Michael Neumann, Ronald J. Stern. Wiley, 1989, xix + 167 pp, $36.95. [ISBN: 0-471-62074-2]

Group Theory, S*, C. *Exploring Small Groups: A Tool for Learning Abstract Algebra.* (IBM PC Software.) Ladnor D. Geissinger. Harcourt Brace Jovanovich, 1989; *User's Manual,* vi + 31 pp, (P). [ISBN: 0-15-526011-1] Award-winning package providing numerous examples of operation tables and groups with on-screen questions (about e.g., order, commuting, elements, subgroups, conjugates, commutativity, associativity, inverses). New tables can be created randomly or by typing. Information generated by the programs is saved for display and review in various screens. The *User's Manual* contains a series of seventy substantial exercises to help guide a student through the exploration of a full complement of properties. LAS

Algebra, T(15-17: 1, 2), L. *Fundamental Concepts of Abstract Algebra.* Gertrude Ehrlich. PWS-Kent, 1991, xi + 340 pp, $38. [ISBN: 0-534-92455-7] Written to serve as a text for either one or two semesters, there is some substantial classical algebra here. Begins with a brief, helpful history of alge-

bra and some set theory. Chapter on group theory includes structure of finite Abelian groups and Sylow theorems. Chapter 3 starts with rings including principal ideal domains and Euclidean domains, concludes with modules and linear algebra. Concluding chapter treats fields, Galois theory, solvability by radicals, and geometric constructions. Numerous exercises (no answers), references, index. JS

Algebra, T(13: 1). *Investigations in Algebra.* Albert Cuoco. MIT Pr, 1990, 623 pp, $29.95. [ISBN: 0-262-53071-6] Logo-based experiments in functions, mathematical induction, combinatorics, and number theory for students who have a background in intermediate algebra and a basic knowledge of Logo programming. "Investigations" is the key word: the author considers the exploration process much more important than the content. Suitable for bright high school or beginning college students as an alternative to calculus. MW

Calculus, T(13-14: 1, 2), S, L. *Explorations in Calculus with a Computer Algebra System.* Donald B. Small, John M. Hosack. McGraw-Hill, 1990, xiv + 226 pp, (P). [ISBN: 0-07-058267-X] A collection of about forty brief, computer-oriented units, each with open-ended exercises, on topics in elementary calculus. Topics are chosen to illustrate the power of "computer algebra systems" for graphical, numerical, and symbolic calculations. The CAS *Maple* is used in most examples, but the book is *not* tied to any one system: "translation" tables are given for parallel commands in *Maple, Mathematica, Macsyma,* and *Derive.* In addition, an introductory chapter treats CAS's in general, and suggests a rationale for their use in undergraduate mathematics. PZ

Calculus, S(13), C. *A Guide to Calculus T/L: A Program for Doing, Teaching, and Learning Calculus.* J. Douglas Child. Macintosh Software and User Guide. Brooks/Cole, 1990, x + 181 pp, (P). [ISBN: 0-534-11817-8] A user-friendly graphical interface to the symbolic mathematical package *Maple* designed to enable calculus students to explore standard topics (functions, graphs, derivatives, integrals, limits, sequences, series) on an electronic whiteboard where objects can be written, moved, combined, and manipulated much as on a sheet of paper. Commands (e.g., `Define f by f(x) = sin x`) are expressed in full, understandable English, displayed for students to choose—so commands don't have to be remembered or typed precisely. Graphs (of functions of one variable only) are easy, as are zooms and scaling. An "authoring" environment enables instructors to ptepare scripts to be redisplayed and explored in class or labs. By clicking "Talk to Maple" one can pre-empt Calculus T/L and interact directly with the underlying *Maple* symbolic system. Requires a Mac Plus or later model, at least 1M of memory, and a hard disk. LAS

Calculus, T(13: 3). *Calculus, Fifth Edition.* Earl W. Swokowski. PWS-Kent, 1991, xxiii + 1159 pp, $48, [ISBN: 0-534-92492-1]; *Calculus of a Single Variable,* 1991, xxi + 765 pp. [ISBN: 0-534-92584-7] A revision of *Calculus with Analytic Geometry, Alternate Edition;* this edition expands discussions, includes guidelines for solving problems, and adds applications to the design of computers and climatic effects. Attractively redesigned with new graphics, some changes in presentation, and some new applications; this remains a traditional calculus text. *Calculus of a Single Variable* consists of the first thirteen chapters of *Calculus, Fifth Edition.* (*First Edition,* TR, December 1975; *Second Edition,* TR, June-July 1979; *Third Edition,* TR, November 1984; *Fourth Edition,* TR, December 1988.) JNC

Real Analysis, T(16-17: 2), L. *Modern Analysis: Measure Theory and Functional Analysis with Applications.* William H. Ruckle. PWS-Kent, 1991, xv + 265 pp. [ISBN: 0-534-92164-7] Intended for a first-year analysis course for graduate students. Chapter one is a review of set theory, the real line, and linear algebra. The other chapters concern metric spaces, measure and integration, functional analysis (e.g., Hahn-Banach Theorem), function and sequence spaces, and applications (e.g., Picard's Theorem, ergodic theory). MLR

Real Analysis, P. *Lecture Notes in Mathematics-1446: Methods of Nonconvex Analysis.* Ed: A. Cellina. Springer-Verlag, 1990, v + 206 pp, $24 (P). [ISBN: 0-387-53120-3] Lectures given at the first session of the Centro Internazionale Matematico Estivo (CIME) held at Varenna, Italy, June 1989. Contains seven lectures. MLR

Real Analysis, P. *Lectures on Singular Integral Operators.* Michael Christ. CBMS Reg. Conf. Ser. in Math., No. 77. AMS, 1990, ix + 132 pp, $27 (P). [ISBN: 0-8218-0728-5] Expansions of lectures given at the NSF-CBMS conference on singular integral operators held at the University of Montana, August 1989. Topics include the Cauchy integral on Lipschitz curves and the T(1) theorem. MLR

Complex Analysis, T(16-17: 1, 2), S, P, L.** *Complex Variables, Second Edition.* Stephen D. Fisher. Math. Ser. Wadsworth, 1990, xiv + 427 pp, $49.95. [ISBN: 0-534-13260-X] A light revision of the *First Edition* (TR, December 1986). An attractive, readable, and enthusiastic—yet rigorous—introduction to theory, methods, and applications of complex analysis in one variable. Although formal prerequisites (multivariable calculus, mainly) are few, exposition is at advanced undergraduate or beginning graduate level. Excellent, varied problem sets will challenge anyone, and point to a wealth of related topics. Many figures support an unusually concrete, geometric viewpoint. Most changes from previous edition are local—rewriting for clarity, new exercises, etc. One more substantial change: Cauchy's theorem, assuming continuity of derivatives, is obtained more directly than formerly using Green's theorem. PZ

Differential Equations, T(18: 1), S, P. *Sturm-Liouville and Dirac Operators.* B.M. Levitan, I.S. Sargsjan. Math. & Its Applic., V. 59. Kluwer Academic, 1991, xi + 350 pp, $149. [ISBN: 0-7923-0992-

8] Parts I and II are parallel developments of the theory and some applications for Sturm-Liouville and one-dimensional Dirac operators. The pattern is similar, proceeding through the spectral theory for the regular and singular cases, the spectrum, distribution of eigenvalues, and inverse problems. References, index. JS

Differential Equations, P. *Theory of Impulsive Differential Equations.* V. Lakshmikantham, D.D. Bainov, P.S. Simeonov. Ser in Modern Appl. Math., V. 6. World Scientific, 1989, x + 273 pp, $58. [ISBN: 9971-50-970-9]

Partial Differential Equations, P. *Stochastic Evolution Systems: Linear Theory and Applications to Non-linear Filtering.* B.L. Rozovskii. Math. & Its Applic., V. 35. Kluwer Academic, 1990, xviii + 315 pp, $129. [ISBN: 0-7923-0037-8] A comprehensive treatment of stochastic partial differential equations—a combination of differential equations and probability theory—with applications in physics, chemistry, and control theory. BC

Partial Differential Equations, T(17: 1, 2). *Principles and Techniques of Applied Mathematics.* Bernard Friedman. Dover, 1990, ix + 315 pp, $8.95 (P). [ISBN: 0-486-66444-9] Unabridged and unaltered republication of work first published by John Wiley in 1956 on how the study of abstract systems can help in the solution of concrete problems. For a one-year graduate course; prerequisites include linear algebra and complex integration. Emphasis on techniques derivable from abstract theory, but stress is on the ideas and not upon a rigorous development. JDEK

Numerical Analysis, S(18), P, L. *Solving Linear Systems on Vector and Shared Memory Computers.* Jack J. Dongarra, *et al.* SIAM, 1991, x + 256 pp, $19.75 (P). [ISBN: 0-89871-270-X] Written by algebraists, not computer scientists, with applied mathematicians in mind. Includes basic aspects of vector and parallel processing, overview of current high-performance computers, and some aspects of implementation and performance, but emphasis is on algorithms and software for dense coefficient systems and direct and iterative methods for sparse ones. Block algorithms prevail. Useful glossary and extensive bibliography. DFA

Operator Theory, T(18: 2), S, P. *A Primer on Spectral Theory.* Bernard Aupetit. Universitext. Springer-Verlag, 1991, xii + 193 pp, $29.80 (P). [ISBN: 0-387-97390-7] A (short) introduction to the techniques of subharmonic functions and analytic multifunctions in spectral theory. Assumes background in subharmonic functions, capacity, and functions of several complex variables. A friendly book. MLR

Functional Analysis, T(17: 2), S, L. *Ergodic Theory.* Karl Petersen. Stud. in Adv. Math., V. 2. Cambridge Univ Pr, 1989, xi + 329 pp, (P). [ISBN: 0-521-38997-6] The author's point of view is that "ergodic theory consists of examples, convergence theorems, the study of various recurrence properties, and the theory of entropy." Each is treated at a basic and advanced level. Assumes measure theory and functional analysis. Many exercises and a sizable bibliography. (1983 hardcover edition, TR, December 1984.) TAV

Functional Analysis, P. *Algebraic Ideas in Ergodic Theory.* Klaus Schmidt. CBMS Reg. Conf. Ser. in Math., No. 76. AMS, 1990, v + 94 pp, $30 (P). [ISBN: 0-8218-0727-7] Presents the ideas that were the basis for the 1989 CBMS Conference at Seattle. Two topics are developed: the influence of operator algebras on dynamics, with some emphasis on Markov shifts, and higher dimensional Markov shifts where little structure exists. TAV

Functional Analysis, T*(16-17: 2), L*. *Functional Analysis and Linear Operator Theory.* Carl L. DeVito. Addison-Wesley, 1990, x + 358 pp, $45.25. [ISBN: 0-201-11941-2] Beautifully motivated elementary introduction to basics of functional analysis. Topics include compact operators, spectral theorems, unbounded operators, and non-normalizable eigenvectors. Good selection of problems with selected solutions at end. A very nice book! BH

Functional Analysis, L. *Approximation Theory and Functional Analysis.* Ed: Charles K. Chui. Academic Pr, 1991, x + 247 pp, $49.95. [ISBN: 0-12-174583-X] Compilation of papers to honor George G. Lorentz on his 80th birthday. Eleven invited papers (mainly survey) on approximation theory and functional analysis presented at a two-day conference at Texas A & M in February 1990. Includes a brief autobiography, lists of Lorentz's publications, Lorentz's doctoral students, and participants. JDEK

Analysis, S(18), P, L. *Topics in Metric Fixed Point Theory.* Kazimierz Goebel, W.A. Kirk. Stud. in Adv. Math., V. 28. Cambridge Univ Pr, 1990, viii + 244 pp, $49.50. [ISBN: 0-521-38289-0] Assumes basic knowledge of analysis and topology (e.g., Banach spaces are not defined but provide the usual setting for the theory presented). First two-thirds of the text deals with classical theory of nonexpansive mappings, last one-third deals with variety of more contemporary topics. Extensive bibliography. BH

Analysis, P. *Computer Aided Proofs in Analysis.* Eds: Kenneth R. Meyer, Dieter S. Schmidt. Math. & Its Applic., V. 28. Springer-Verlag, 1991, xvi + 251 pp, $32. [ISBN: 0-387-97426-1] Twenty-one papers from a 1989 conference at the University of Cincinnati. BC

Analysis, P. *Recent Advances in Fourier Analysis and Its Applications.* Eds: J.S. Byrnes, Jennifer L. Byrnes. NATO ASI Ser. C, V. 315. Kluwer Academic, 1990, x + 686 pp, $174. [ISBN: 0-7923-0875-1] Forty-one papers from a 1989 conference. Applications run the gamut from X-ray crystallography to neurocomputer architectures. Special attention to polynomials with restricted (e.g., unimodular) coefficients. BC

Analysis, T(14-17). *Real Analysis.* Norman B. Haaser, Joseph A. Sullivan. Dover, 1991, ix + 341 pp, $8.95 (P). [ISBN: 0-486-66509-7] Intended to

provide a smooth transition from calculus to abstract analysis. Covers metric spaces, normed linear spaces, inner product spaces, and develops the Lebesgue integral. Also includes topics involving approximation theory, the Banach fixed point theorem and its applications, and Stieltjes integrals. (1971 Van Nostrand Reinhold edition, TR, August-September 1971.) KS

Algebraic Geometry, T(17-18: 1), S, P, L. *Algebraic Geometry for Scientists and Engineers.* Shreeram S. Abhyankar. Math. Surveys & Mono., No. 35. AMS, 1990, xiii + 295 pp, $87. [ISBN: 0-8218-1535-0] From the Preface: "I am simply resurrecting the concrete and ancient methods of Shreedharacharya (500 A.D.), Bhaskaracharya (1150 A.D.), Newton (1660), Sylvester (1840), Salmon (1852), Max Noether (1870), Kronecker (1882), Cayley (1887), and so on." Not a bad bunch to bring back. BC

Differential Geometry, P. *Lecture Notes in Mathematics-1451: Global Geometry and Mathematical Physics.* L. Alvarez-Gaumé, *et al.* Springer-Verlag, 1990, ix + 197 pp, $24 (P). [ISBN: 0-387-53286-2] Proceedings of a 1988 conference in Montecatini Terme, Italy. Includes papers on moduli spaces, conformal field theory and string theory, and the geometry of the KP equation. JO

Differential Geometry, T?(17-18: 1), S, P. *Introduction to Differential Geometry for Engineers.* Brian F. Doolin, Clyde F. Martin. Pure & Appl. Math., V. 136. Marcel Dekker, 1990, xii + 163 pp, $45. [ISBN: 0-8247-8396-4] Written to provide control engineers with the necessary background and vocabulary in differential geometry to read research papers in geometric and nonlinear control. Topics include manifolds, tangent spaces, vector fields, differential forms, Lie groups, and Grassmannian techniques for control theory. Some background in differential geometry and topology is assumed since terms like diffeomorphism and second countable Hausdorff are used, but not defined. Only ten exercises; references. OJ

Differential Geometry, T(17-18), P. *Riemannian Geometry, Second Edition.* Sylvestre Gallot, Dominique Hulin, Jacques Lafontaine. Universitext. Springer-Verlag, 1990, xiii + 284 pp, $29 (P). [ISBN: 0-387-52401-0] This edition adds sections on surfaces with constant negative curvature and on conformal geometry, and makes some corrections. (*First Edition,* TR, June-July 1988.) OJ

Geometry, T(18: 2), S, P. *Lecture Notes in Mathematics-1441: Géométrie et théorie des groupes.* M. Coornaert, T. Delzant, A. Papadopoulos. Springer-Verlag, 1990, x + 165 pp, $18 (P). [ISBN: 0-387-52977-2] Based largely on Gromov's work on hyperbolicity, especially his 1988 paper "Hyperbolic Groups." Discusses hyperbolic metric spaces and hyperbolic groups. Other subjects include quasi-geodesics, isoperimetric inequalities, and isometries. Reviews basic facts of automata theory, and shows that hyperbolic groups are automatic. In French. AD

Geometry, T*(15: 1, 2). *Geometry: A Metric Approach with Models, Second Edition.* Richard S. Millman, George D. Parker. Undergrad. Texts in Math. Springer-Verlag, 1991, xiii + 370 pp, $39. [ISBN: 0-387-97412-1] Revised edition of work first published in 1981 (TR, May 1983). Topics covered are neutral geometry, theory of parallels, hyperbolic geometry, classical Euclidean geometry, Bolyai approach to theory of area, and isometries. Metric approach. Emphasis on illustrating definitions and axioms with models. Added material in this edition includes some expository exercises. Basic prerequisite is mathematical maturity; calculus helpful. Currently available (from the second author): a program written in Pascal which allows graphical explorations and calculations in the Poincaré plane. Authors believe that mathematics is learned by doing and not by reading. JDEK

Geometry, S, C, P. *Designer Fractal: Mathematics for the 21st Century.* (Macintosh Software.) Peter Van Roy, Linda Garcia, Bernt Wahl. Dynamic Software (PO Box 7534, Santa Cruz, CA 95061), 1988. *User Manual,* 15 pp, (P). A design tool called FractaSketch that creates, saves, and prints fractal designs by replication of a template. Includes tools to create designs at several different levels of iteration; to magnify, reduce, and relocate the design once created; to save in the Clipboard or Scrapbook; to print on a laser printer at good resolution; and to create new templates. Calculates and displays the fractal dimension of each pattern. Includes several dozen prepared templates to produce innumerable designs, including all classic fractals. LAS

Geometry, S, C, P. *The Desktop Fractal Design System.* (Macintosh Software.) Michael F. Barnsley. Academic Pr, 1990, *Handbook,* v + 40 pp, $39.95 (P) set. [ISBN: 0-12-079064-5] An implementation of the Iterated Function System (IFS) for encoding fractal-generating affine transformations into efficient codes (typically, a few lines, each containing six digits) from which the "chaos algorithm" can reproduce fractal pictures of arbitrary complexity. Software displays these fractal images, permiting dynamic modification of IFS code to make changes in images visible. Also contains an implementation of the escape-time algorithm that computes attractor sets, Julia sets, and the Mandelbrot set. *Handbook* explains not only how to use software but also how IFS codes generate fractals, and how the two key algorithms can be implemented in simple Basic programs. A useful supplement to Barnsley's 1988 *Fractals Everywhere* (TR, April 1989; Extended Review, March 1990). Software fails to run on some newer Macintosh models. LAS

Geometry, S(9-13). *The Theorem of Pythagoras.* Tom M. Apostol. Project Mathematics! California Institute of Technology, 1988, 22 minute videotape; *Program Guide and Workbook,* 30 pp, (P). A motivational and instructional video introduction to right triangles: opens with typical survey problems, covers standard algebraic formulation (includ-

ing Pythagorean triples), and then gives three visual proofs (Chinese, Euclid, dissection). Concludes with extensions to other shapes and dimensions. Intended for middle school students. Accompanying *Workbook* contains discussion, extensions, and exercises. Available from state Departments of Education, from MAA, and from the Cal Tech bookstore. LAS

Algebraic Topology, T(17-18), S, P. *Modern Geometry—Methods and Applications, Part III: Introduction to Homology Theory.* B.A. Dubrovin, A.T. Fomenko, S.P. Novikov. Transl: Robert G. Burns. Grad. Texts in Math., V. 124. Springer-Verlag, 1990, ix + 416 pp, $59. [ISBN: 0-387-97271-4] Develops singular homology and cohomology theory together. Emphasizes geometric aspects and avoids homological algebra where possible. Also presents de Rham and sheaf cohomologies. Wide range of topics, e.g., classification of surfaces, homotopy groups of spheres, Bott periodicity. Includes chapters on Morse theory and cobordism. Assumes background from Part II on smooth manifolds, homotopy theory, and fiber spaces. Few exercises; bibliography. OJ

Topology, T(18), S, P. *Topological Fields and Near Valuations.* Niel Shell. Pure & Appl. Math., V. 135. Marcel Dekker, 1990, xiii + 233 pp, $89.75. [ISBN: 0-8247-8412-X] A survey of the theory of topological fields which introduces the notion of a prevaluation as a tool to study locally bounded topologies. Requires a solid background in algebra and point set topology. Does not include exercises. CEC

Topology, T?(17), P, L. *Introduction to Uniform Spaces.* I.M. James. London Math. Soc. Lect. Note Ser., V. 144. Cambridge Univ Pr, 1990, 148 pp, $24.95 (P). [ISBN: 0-521-38620-9] A uniform structure on a set X is a collection Ω of relations on X such that 1) Ω is a filter on $X \times X$; 2) each $D \in \Omega$ contains the diagonal of $X \times X$; 3) $D \in \Omega$ implies $D^{-1} \in \Omega$; 4) $D \in \Omega$ implies $E \circ E \subset D$ for some $E \in \Omega$. Topological spaces are generalizations of uniform spaces. The author examines the relationship between the two and also develops the theory of uniform transformation groups. Forty exercises at end of book. Rather dry presentation. BH

Operations Research, T(15-16: 1). *Introduction to Linear Programming: Applications and Extensions.* Richard B. Darst. Pure & Appl. Math., V. 141. Marcel Dekker, 1991, xii + 353 pp. [ISBN: 0-8247-8383-2] An elementary introduction to the theory and practice of solving linear programming problems. Also covers techniques for some closely related problems including quadratic programming, minimization of quadratic functions, and a variety of network problems. AO

Operations Research, T(14-15), L. *Linear Programming and Applications: A Course Text.* Will McLewin. Input-Output (3 Wyndham Place, London), 1990, xiv + 216 pp, $33. [ISBN:0-904870-11-1] An elementary linear programming text. Uses both tableau-based and matrix-based approaches. Contains the standard coverage of the simplex method, duality, sensitivity analysis, and the usual applications to network, assignment, and game theory problems. Discusses Khachian's Ellipsoid Method, but does not mention Karmarkar's technique. Limited number of exercises, mostly theoretical in nature. (1981 edition, TR, August-September 1982.) SM

Operations Research, T(18: 1), P, L. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Anthony V. Fiacco, Garth P. McCormick. Classics in Appl. Math., V. 4. SIAM, 1990, xvi + 210 pp, $26.50 (P). [ISBN: 0-89871-254-8] A corrected, unabridged republication of the original classic, discussing techniques of solving constrained nonlinear programming problems using a sequence of unconstrained problems. Very well-written; part of a very important series published by SIAM. (1968 John Wiley edition, TR, May 1969.) SM

Operations Research, T(14-15: 1), S, L. *Mathematics Applied: An Introduction to Mathematical Modeling.* William J. Adams. Pace & Pace, 1990, viii + 215 pp, $15.75 (P). An undergraduate textbook with minimal prerequisites. A semester of calculus and some knowledge of deductive logic is sufficient. Four approaches are covered: linear programming and random processes (the bulk of the book), geometry and differential equations. Plenty of exercises, some with solutions. Inexpensive. SM

Dynamical Systems, $\mathcal{E}$*(12-16). *The Beauty and Complexity of the Mandelbrot Set.* John H. Hubbard. Science Television (Box 2498, Times Square Station, New York, NY 10108), 1989. *School Edition,* 47 minute videotape, $49, [ISBN: 1-878310-03-8]; *University Edition,* 73 minute videotape, $59. [ISBN: 1-878310-02-x] A polished, well-motivated lecture by John Hubbard, beginning with Newton's method applied to quadratic iterations from a high school student's perspective and concluding with a full exploration of Julia sets and the Mandelbrot's set. Hubbard lectures "the old fashioned way," effectively and elegantly, using a whiteboard to draw suggestive pictures that help the viewer interpret many beautiful still and moving color graphics. Computer methods are integrated very effectively: mouse movements demonstrate how different seed points create different orbits; an incredible zoom reveals the iterated complexity of the Mandelbrot set. Brief concluding section raises issues of complexity in analogy with DNA, offering Mandelbrot's theory as a "message of hope for biology" that the genetic code may not be as complex as it now appears. The longer *University Edition* contains an extensive section on equipotential lines, field lines, and fractional labels as a means of explaining in precise terms the relation between Julia sets and the Mandelbrot set. The incredible richness of ideas conveyed in these tapes would require repeated viewings (with interspersed study) for full appreciation. LAS

Dynamical Systems, S, C. *Chaos: Mathematics for the 21st Century.* (Macintosh Software.) Bernt Wahl. Aerial Pr, 1989. A series of programs

that display graphs of various dynamical processes (Lorenz attractor, logistic map, Henon maps, pendulum, Van der Pol, predator-prey) with various options for changing parameters, both in the underlying equations and in the form of the graphical representation. Each process starts with default parameter values; other recommended values are available via buttons, as are user-provided parameters. LAS

Dynamical Systems, S. *Fractals on Hewlett-Packard Supercalculators, Preliminary Edition.* Yves Nievergelt. New Directions in Math Ser. Wadsworth, 1991, vi + 74 pp, $8 (P). [ISBN: 0-534-15037-3] A workbook on fractals (Julia sets, von Koch's snowflake, Hausdorff dimension) with exercises (and sample programs) for HP series of programmable calculators (HP-28C, HP-28S, HP-48SX). Concludes with exploratory term projects. Provides mathematical exposition for students at all levels—beginning with those who do not yet know complex numbers. Published without binding or cover to keep the price down! LAS

Dynamical Systems, S(12-16), P*.** *Fractals: An Animated Discussion.* Heinz-Otto Peitgen, *et al.* WH Freeman, 1990, 63 minute videotape, $59.95 (Individual); $149.95 (Institutional). *Documentation,* 20 pp. [ISBN: 0-7167-2213-5] Interviews with Edward Lorenz and Benoit Mandelbrot set the stage for a wide-ranging visual and scientific introduction to fractal behavior (coast lines, self-similarity, butterfly effect). Central exemplar is a pendulum swinging over three magnets illustrated via videotape and computer simulation. Effective historical and contextual presentation; little direct mathematics (although much is given, with references, in the brief accompanying *Documentation*). Mandelbrot discusses the origin of his discoveries, his philosophy of geometry and applied mathematics, and his perspective on mainstream mathematics of his generation. Fabulous films illustrate the relation between Julia sets and the Mandelbrot set; visual landscapes blend artificial worlds with real footage; even the music is synthesized and composed from fractral algorithms. A true delight for the eye, for the mind, and for the mind's eye. LAS

Dynamical Systems, S, C*, P*.** *The Beauty of Fractals Lab.* (Macintosh Software.) Hartmut Jürgens, Heinz-Otto Peitgen, Dietmar Saupe. Macintosh Software by Thomas Eberhardt and Marc Parmet. Springer-Verlag, 1990, $49. [ISBN: 0-387-14205-3] A toolkit to construct full color Julia sets and the Mandelbrot set in 2, 2.5, or 3-dimensional perspectives using a variety of algorithms (some for speed, some for accuracy). Supports multiple windows with different images; full control over number of iterations, color (or gray-scale) mappings, distance estimates, image size, and light sources. Cursor tools (in a palette on each window) permit insertion of new seeds, blow-ups of selected regions, construction of Julia sets corresponding to points selected from the Mandelbrot set, display of first orbit steps of selected initial values, and visual display of distance estimates. Images can be saved in Scrapbook for importation into other graphics or word processing packages. Includes several interesting demos of classic images; requires a Mac II with at least 1 megabyte of memory running system 6.0.2 or later. LAS

Control Theory, T(18), P, L. *Decentralized Control of Complex Systems.* Dragoslav D. Šiljak. Math. in Sci. & Eng., V. 184. Academic Pr, 1991, xiv + 527 pp, $74.95. [ISBN: 0-12-643430-1] Discusses the control of large scale systems by a divide-and-conquer strategy. The central complex system is partitioned into subsystems, each of which is subject to a less complex controller. The hope is to apply the ideas of distributed parallel processing to complex control problems. SM

Probability, T(17-18: 2), S. *Probability: Theory and Examples.* Richard Durrett. Wadsworth, 1991, x + 453 pp, $53.95. [ISBN: 0-534-13206-5] Intended for a year course for students familiar with measure theory. In addition to the usual topics and numerous examples, the author includes material on stochastic processes including Markov chains on general state space, renewal theory, and central limit theorems for martingales. TAV

Probability, T(14-15). *A Primer in Probability, Second Edition, Revised and Expanded.* Kathleen Subrahmaniam. Stat.: Textbooks & Mono., V. 111. Marcel Dekker, 1990, xi + 318 pp, $45. [ISBN: 0-8247-8348-4] Undergraduate, post-calculus course in probability. Includes introduction to statistical inference (simple testing theory). Numerous examples and exercises as well as a discussion of various interpretations of probability and its role in scientific investigations. Includes standard topics. (*First Edition,* TR, March 1980.) MK

Probability, P. *Lecture Notes in Mathematics-1442: Quantum Probability and Applications V.* Eds: L. Accardi, W. von Waldenfels. Springer-Verlag, 1990, vi + 413 pp, $45 (P). [ISBN: 0-387-53026-6] Proceedings of a 1988 workshop held in Heidelberg. Thirty short papers on miscellaneous topics. JO

Stochastic Processes, T(18: 2), P. *Stochastic Flows and Stochastic Differential Equations.* Hiroshi Kunita. Stud. in Adv. Math., V. 24. Cambridge Univ Pr, 1990, xiv + 346 pp, $69.50. [ISBN: 0-521-35050-6] The author attempts to develop the relationship between stochastic flows of diffeomorphisms and stochastic differential equations, using the latter to study the properties of the flows. Contains many examples and exercises, and a large bibliography. TAV

Stochastic Processes, P. *Stochastic Integrals: An Introduction.* Heinrich von Weizsäcker, Gerhard Winkler. Adv. Lect. in Math. Friedr Vieweg & Sohn, 1990, ix + 332 pp, DM 78 (P). [ISBN: 3-528-06310-6] A nicely written introduction to stochastic integrals and the Ito calculus for semi-martingales. The material is deep and the reader should expect to spend effort grasping the details, but the subject is fundamental in modern stochastic theory. TAV

Stochastic Processes, P. *Dependence with Com-*

plete Connections and Its Applications. Marius Iosifescu, Serban Grigorescu. Tracts In Math., V. 96. Cambridge Univ Pr, 1990, xiv + 304 pp, $69.50. [ISBN: 0-521-33331-8] Dependence with complete connection is a generalization of Markovian dependence. The book presents a theoretical development and several applications. Contains appendices to assist the readers with prerequisites and a very extensive bibliography. TAV

Stochastic Processes, T(18: 2), P, L. *Stationary Stochastic Models.* Andreas Brandt, Peter Franken, Bernd Lisek. Wiley, 1990, 344 pp, $89.95. [ISBN: 0-471-92132-7] Considers both discrete and continuous time models. Emphasizes queueing theory in the last half of the book. Tersely written in textbook style. No exercises. Recommended for readers with some prior experience in the field. Extensive references. SM

Stochastic Processes, P. *Stochastic Processes and their Applications in Mathematics and Physics.* Eds: Sergio Albeverio, Philippe Blanchard, Ludwig Streit. Math. & Its Applic., V. 53. Kluwer Academic, 1990, xiii + 402 pp, $129. [ISBN: 0-7923-0894-8] Twenty-two papers—revised versions of papers given in 1985 at Bielefeld. On three general areas: geometric structure of diffusion processes, probabilistic problems from quantum physics, and the interaction between probability theory and statistical mechanics. TAV

Stochastic Processes, P. *Lecture Notes in Mathematics-1444: Stochastic Analysis and Related Topics II.* Eds: H. Korezlioglu, A.S. Ustunel. Springer-Verlag, 1990, 268 pp, $29 (P). [ISBN: 0-387-53064-9] Proceedings of a workshop in Silivri, Turkey, July 1988. Major papers on asymptotic problems in Weiner integration and applications of anticipating stochastic calculus to stochastic differential equations. TAV

Stochastic Processes, P. *Limit Theorems on Large Deviations for Markov Stochastic Processes.* A.D. Wentzell. Math. & Its Applic., V. 38. Kluwer Academic, 1990, xv + 176 pp, $79. [ISBN: 0-7923-0143-9] A clean translation of the 1986 Russian original edition. After preliminaries, the author investigates the action functional for Markov processes, precise asymptotics, and probabilities of large deviations. TAV

Elementary Statistics, T(13-14: 2). *Statistics for Business and Economics.* Debra Olson Oltman, James R. Lackritz. Brooks/Cole, 1991, xxii + 984 pp, $45. [ISBN: 0-534-14430-6] Introduction to descriptive and inferential statistics with some probability. Chapters on ANOVA, regression, nonparametric statistics, time series, index numbers, and decision making. Includes MINITAB as well as other computer output. Real danger students will lose sight of important concepts because of the encyclopedic listing of terminology and formulas. Some of the figures and tables are extremely difficult to read. RWJ

Computational Statistics, T(16), S, P. *Modern Statistical, Systems, and GPSS Simulation: The First Course.* Zaven A. Karian, Edward J. Dudewicz. WH Freeman, 1991, xvi + 468 pp, $56.95. [ISBN: 0-7167-8232-4] Introduces theory and implementation of discrete-event simulation. Addresses theoretical basis for simulation, how long to run a simulation, and how to analyze output for experimental goals and integrates these with GPSS, a simulation language. Written at an undergraduate probability level and contains a variety of examples from business, time-shared computer analysis, and emergency medical systems. Handy reference for standard simulation techniques. MK

Computational Statistics, S(13-16), C. *Statistics with Stata.* Lawrence C. Hamilton. Brooks Cole, 1990, vii + 171 pp, $31.95 (P). [ISBN: 0-534-12850-5] Supplementary text for using STATA in classroom setting. Runs on MS-DOS and UNIX platforms. Graphics capabilities are excellent for intended level. Although not a do-all package, covers a lot of ground: EDA methods, ANOVA, regression, logistic and robust regression, diagnostics (including Cook's distance, DFFITS, VIF, added-variable plots, etc.), survival analysis, non-parametrics, time series, and more. Easy for students to learn; nice for comparison with more elaborate, research-level packages. MK

Statistics, T(18), P. *Order Statistics and Inference: Estimation Methods.* N. Balakrishnan, A. Clifford Cohen. Stat. Model. & Decision Sci. Academic Pr, 1991, xix + 377 pp, $79.95. [ISBN: 0-12-076948-4] Consolidates developments in mathematical methods based on order statistics. Full discussion of moment and maximum-likelihood estimation, with emphasis on linear unbiased and optimal linear estimation based on complete or censored samples. Emphasis is on theory although nearly all discussion is illustrated with brief examples. Includes an encyclopedic bibliography. MK

Statistics, P. *Monotonicity Properties of Infinitely Divisible Distributions.* B.G. Hansen. CWI Tract, V. 69. Mathematisch Centrum, 1990, ii + 101 pp, Dfl. 38.50 (P). [ISBN: 90-6196-380-X] "Studies the relationship between an infinitely divisible distribution (i.e., solution to a generalized central limit problem) and its Lévy spectral function." BH

Statistics, T(14-17: 1, 2). *Statistics for Research, Second Edition.* Shirley Dowdy, Stanley Wearden. Ser. in Prob. & Math. Stat. Wiley, 1991, xvii + 629 pp, $49.95. [ISBN: 0-471-85703-3] Revision of the authors' 1983 *First Edition* (TR, January 1984). Significant change is the addition of sections on computer usage in each chapter, with techniques illustrated by SAS commands and output. Another new feature is the inclusion of nonparametric procedures as analogs to previously discussed parametric tests. The chapter on multiple regression has also been rewritten to emphasize modern computer output. RSK

Statistics, T(18: 1), P, L. *Nonparametric Function Estimation, Modeling, and Simulation.* James R. Thompson, Richard A. Tapia. SIAM, 1990, xvi

+ 304 pp, $32.50 (P). [ISBN: 0-89871-261-0] Gives a survey of techniques for one-dimensional nonparametric density estimation. The harder problem of multi-dimensional estimation is also addressed. Model building and mathematical optimization are also each given one chapter. No exercises. SM

Programming, S(14-16), P, L. *A C++ Toolkit.* Jonathan Shapiro. Ser. in Innovative Technology. Prentice-Hall, 1991, xiii + 231 pp, (P). [ISBN: 0-13-127663-8] A handbook for developing C++ object classes that may be used and reused in applications. The author illustrated principles for writing reusable classes with his own example classes for standard data structures. Introduction, including a reader's guide to C++; presentation of example classes; storage management and performance tuning; code listings. Thought-oriented exercises. RB

Programming, T(13-14), S, C, P, L. *Programming in MacScheme.* (Macintosh Software.) Michael Eisenberg, *et al.* MIT Pr, 1990, xviii + 388 pp, $32.50 (P). [ISBN: 0-262-55018-0]; *MacScheme Manual and Software.* MIT Pr, 1990, 194 pp, $37.50 (P). [ISBN: 0-262-62077-4] A manual, a Macintosh student disk, and an introductory text for Scheme, a simple, learnable dialect of Lisp that was developed in part to provide a basis for teaching introductory computer science. Text spirals among projects, language structure, and debugging; final few chapters emphasize procedures as first class objects—the "mantra" of Scheme. Appendices discuss the MacScheme interface, offer solutions to selected exercises, references, and several indices. The *Manual* begins with a brief overview of the language, the compiler, and the debugger; the major section is a detailed reference manual for the language. Appendices discuss "Easy Graphics," an extension built into MacScheme to take advantage of the Macintosh graphics capability. Includes a complete list of error messages, and other reference lists. The MacScheme disk includes several sample programs and special procedures required to make MacScheme compatible with other versions (e.g., Abelson & Sussman's well-known text *Structure and Interpretation of Computer Programs*). Requires a Mac Plus, SE, or II with at least 1M memory. LAS

Programming, S(14), P. *Designing Screen Interfaces in C.* James L. Pinson. Yourdon Pr (Distr: Prentice-Hall), 1991, xvi + 267 pp, (P). [ISBN: 0-13-201583-8] For the experienced C programmer. Shows how to create windowing, pull-down menus, help screens, data input screens, point-and-shoot file selection, and other features to enhance software for IBM PCs and clones. Modular. Includes much highly-commented source code. DFA

Languages, P, L.** *C: A Reference Manual, Third Edition.* Samuel P. Harbison, Guy L. Steele, Jr. Prentice-Hall, 1991, viii + 392 pp, (P). [ISBN: 0-13-110933-2] An *essential* book for serious C programmers. This edition incorporates ANSI C into the main text and provides detailed advice on writing programs that are portable between ANSI C and traditional C compilers. Also new in this edition are a collection of study exercises (with solutions). AO

Languages, P, L. *The Definition of Standard ML.* Robin Milner, Mads Tofte, Robert Harper. MIT Pr, 1990, xi + 101 pp, $15 (P). [ISBN: 0-262-63132-6] The official, formal definition of the programming language ML, a relatively new functional programming language that is strongly typed and also polymorphic. It also provides imperative language constructs, parametric modules, and a sophisticated exception mechanism. AO

Languages, T(15-17: 1), S, L. *Programming Linguistics.* David Gelernter, Suresh Jagannathan. MIT Pr, 1990, xx + 411 pp, $37.50. [ISBN: 0-262-07127-4] Designed to be used as a textbook for courses in the design and evolution of programming languages. Emphasizes the importance of aesthetics and elegance in the design of programming languages. An interesting and innovative text that deserves serious considerations. AO

Algorithms, P. *Proceedings of the Second Annual ACM-SIAM Symposium on Discrete Algorithms.* Chair: Alok Aggarwal, *et al.* SIAM, 1991, 482 pp, $45 (P). [ISBN: 0-89871-271-8] 53 research papers, primarily on graph algorithms and computational geometry, selected for and presented at a discrete algorithms conference, January 1991. JPH

Algorithms, T*(15-16: 1), L. *Algorithms and Data Structures: Design, Correctness, Analysis.* Jeffrey H. Kingston. Intern. Comput. Sci. Ser. Addison-Wesley, 1990, ix + 315 pp. [ISBN: 0-201-41705-7] A textbook for a data structures and algorithm analysis course that emphasizes depth over breadth. Emphasizes the use of abstract data types, proofs of correctness, and appropriate algorithm design techniques. Unusual features of the book include its treatment of amortized complexity analysis, splay trees, and Fibonacci heaps. Example programs are presented in Modula-2. AO

Computer Systems, P. *An Architecture for Combinator Graph Reduction.* Philip John Koopman, Jr. Academic Pr, 1990, xv + 155 pp, $34.95. [ISBN: 0-12-419240-8] "Traditionally, it has been assumed that advanced programming languages [including functional languages] require radically different, non-von Neumann architectures for efficient execution. This book explores mapping functional programming languages onto conventional architectures using a combination of techniques from the fields of computer architecture and implementation of advanced programming languages [in particular, combinator graph reduction]." RB

Computer Systems, C, P. *Leo: A Technical Editor.* (IBM PC Software.) ABK Software (4495 Ottawa Place, Boulder, CO 80303). *User Manual,* 1990, 136 pp, (P). A text editor for IBM PC systems with a graphics display designed as a front end for TEX. Menus and special keys are used to provide special features (font changes, accents, math-mode) which are displayed (or approximated) directly on the screen as they will be when printed. The file

created by Leo is a TEX file, and can be input directly into TEX. Or it can be printed by a pseudo-TEX processor Leopr which approximates most important features of TEX. Leo handles only a subset of TEX in a way that hides the TEX input code from the user. Very thorough *User Manual.* LAS

Computer Systems, C, P. *X Window System, User's Guide, OSF/Motif Edition, Volume Three.* Valerie Quercia, Tim O'Reilly. O'Reilly & Assoc, 1990, xxvi + 709 pp, $30 (P). [ISBN: 0-93175-61-7] Complete manual for the *X Window System, Version 11, Release 4.* Assumes X is installed and running and that you are running UNIX, although the UNIX dependencies are slight. Includes introduction and how to get started, and progresses to discussions of the mwm window manager, font specification, graphics utilities, other clients, and customizing X. MK

Computer Systems, C, P. *X Toolkit Intrinsics Programming Manual: OSF/Motif 1.1 Edition for X11, Release 4, Volume 4.* Adrian Nye, Tim O'Reilly. O'Reilly & Assoc, 1990, xxxiii + 632 pp, $30 (P). [ISBN: 0-937175-62-5] Manual for writing X Window System programs using the Xt Intrinsics Library. Divided into two volumes, this volume (*IV*) provides explanation of the toolkit. Includes tutorial and many task-oriented examples. This is a companion to *Volume V, X Toolkit Reference Manual.* Includes discussion of widget sets, instance hierarchies, creating basic menus and dialog boxes, C graphics models, resource management, interclient communications, and more. MK

Computer Systems, S(13-16). *Introducing Mathematica.* Stephen Wolfram. Science Television (Box 2498, Times Square Station, New York, NY 10108), 1989, 60 minute videotape, $49. [ISBN: 1-878310-01-1] An hour-long videotape of Stephen Wolfram talking (casually and informally) into the camera while typing *Mathematica* at a keyboard. Split screen video shows both Wolfram's lecture and *Mathematica*'s screen display. In three parts: quick overview of numerical, symbolic, and graphical features; detailed examination of each of these features; an introduction to extensions (e.g., programming, application packages, TEX and Fortran outputs). Includes "animation" of surface graphs changing in response to a control parameter. Illustrates all key features of *Mathematica;* excellent introduction for those without sufficient computer experience or resources to use the actual package. LAS

Computer Systems, L*. *sed & awk.* Dale Dougherty. O'Reilly & Assoc, 1990, xvii + 394 pp, $24.95 (P). [ISBN: 0-937175-59-5] A detailed introduction for beginners to two UNIX text processing tools: the stream editor sed and the pattern-recognition language awk (named for its authors Aho, Weinberger, Kernighan). Both tools permit systematic transformations (permutations, substitutions, rearrangements) of strings of words or data, thus serving as useful pre-processors for other packages (e.g., statistical software, text formatters) and as programming tools of their own. Includes two large examples—an interactive spell checker and an indexing program—plus scripts (i.e., programs) submitted by colleagues to illustrate the diverse power of these tools. LAS

Computer Systems, P. *SAA: A Guide to Implementing IBM's Systems Application Architecture.* Jerrold M. Grochow. Yourdon Pr (Distr: Prentice-Hall), xxvi + 385 pp. [ISBN: 0-13-785759-4] IBM introduced SAA in 1987 as a comprehensive framework for defining "IBM compatibility." Computing applications that adhere to SAA provide standard user interfaces, share programming tools and software development strategies, and can communicate among diverse IBM computer systems. This guide presents the architecture for managerial and technical personnel in industry. RB

Computer Science, P. *Realistic Compiler Generation.* Peter Lee. Found. of Computing Ser. MIT Pr, 1989, xv + 246 pp, $32.50. [ISBN: 0-262-12141-7] Presents a survey of semantics-based compiler generation and a new method ("high-level semantics") of describing programming language semantics. In contrast to previous methods, the new method makes the automatic generation of efficient compilers possible. Based on the author's Ph.D. thesis. AO

Computer Science, P. *Object-Oriented Databases with Applications to CASE, Networks, and VLSI CAD.* Eds: Rajiv Gupta, Ellis Horowitz. Prentice-Hall, 1991, xiv + 447 pp. [ISBN: 0-13-629833-8] Object-oriented database systems ("OODBs") have recently appeared in response to conventional relational database system limitations in applications such as computer-aided design (CAD), manufacture (CAM), software engineering (CASE). This collection of twenty-one articles introduces object-oriented concepts, surveys existing OODBs, presents real-world application examples, and includes a brief overview of C++ language and database enhancements of C++. RB

Applications, P. *Proceedings of the Fourth European Conference on Mathematics in Industry.* Eds: Hansjörg Wacker, Walter Zulehner. European Consort. for Math. in Industry, V. 6. Kluwer Academic, 1991, x + 433 pp, $148. [ISBN: 0-7923-1036-5] Proceedings of the Fourth ECMI Conference on Industrial Mathematics held at Strobl in Austria, May 29-June 2, 1989. Articles in following sections: invited lectures (includes models for crystallization of polymers, information technology, trends in higher education, thermal behavior of thyristors with cooling devices), steel processing, chemical engineering, as well as a wide variety of contributed lectures. MK

Applications, P, L*. *The Ubiquity of Chaos.* Ed: Saul Krasner. AAAS, 1990, viii + 247 pp, (P). [ISBN: 0-87168-350-4] Nineteen papers on diverse applications of chaos that have been presented at AAAS annual meetings over the past several years. Topics include brain waves, quantum chaos, business cycles, childhood epidemics, stellar variability, and international security—a stunning array of applications, united in a helpful common index, that documents

the "ubiquity" of chaos. LAS

Applications (Engineering), T(18), P. *Stochastic Finite Elements: A Spectral Approach.* Roger G. Ghanem, Pol D. Spanos. Springer-Verlag, 1991, x + 214 pp, $49.50. [ISBN: 0-387-97456-3] With applications to mechanical engineering in mind, this book extends the well-known method of finite elements to problems involving uncertainty. Primarily theoretical in nature, the book concludes with three examples. SM

Applications (Physical Science), P. *Applied and Industrial Mathematics, Venice-1, 1989.* Ed: Renato Spigler. Math. & Its Applic., V. 56. Kluwer Academic, 1991, xiii + 374 pp, $136. [ISBN: 0-7923-0521-3] Thirty-one papers from a 1989 conference on industrial-strength mathematics including fluid-flow modelling, nonlinear waves, propagation in random media, and transport phenomena. BC

Applications (Physical Science), S(16-17), P, L*. *Crystalline Symmetries: An Informal Mathematical Introduction.* Marjorie Senechal. Adam Hilger, 1990, xi + 137 pp, $39. [ISBN: 0-7503-0041-8] An excellent, up-to-date introduction to the mathematical basis of crystallography, from Platonic solids to Penrose tiles. Ideal undergraduate seminar material. BC

Applications (Physics), S(17-18), P. *Electro-Diffusion of Ions.* Isaak Rubinstein. Stud. in Appl. Math., V. 11. SIAM, 1990, ix + 254 pp, $44.75. [ISBN: 0-89871-245-9] An innovative approach that allows the unified treatment of such diverse topics as electrolyte solutions, ion-exchangers, ion-selective membranes, and semiconductors. This book is intended to attract applied mathematicians to a remarkably under-developed area. MU

Applications (Physics), T(17-18: 1), S, P, L. *Direct and Inverse Problems: Potentials in Quantum Scattering.* Boris N. Zakhariev, Allina A. Suzko. Transl: G. Pontecorvo. Springer-Verlag, 1990, xiii + 223 pp, $32 (P). [ISBN: 0-387-52484-3] With the inclusion of new material concerning exactly solvable models, this fine translation supercedes the Russian original. Part I—One-Dimensional, One-Channel Systems—provides a brief but careful introduction culminating in the Levinson Theorem. Part II is devoted to multi-channel problems. Illustrations, exercises, and a lengthy bibliography are also provided. MU

Applications (Physics), P. *Random Matrices, Revised and Enlarged Second Edition.* Madan Lal Mehta. Academic Pr, 1991, xviii + 562 pp, $99.50. [ISBN: 0-12-488051-7] Random matrices are used in this volume as a tool for statistical quantum mechanics—as a means of computing statistical properties of chaotic systems, which are far more common than the integrable systems studied in ordinary texts. Major focus is on Gaussian ensembles, especially on evidence for a "central limit" conjecture that the statistical properties of a few eigenvalues of a large random matrix are independent of the probability distribution of the elements of the matrix. This conjecture, backed by extensive numerical evidence, helps tie together issues such as nuclear physics, Ising models, chaotic dynamics, and the pattern of zeroes of the Riemannan Zeta function (whose local fluctuation properties appear identical to those of the eigenvalues of matrices from a unitary ensemble. (*First Edition,* TR, March 1968.) LAS

Applications (Physics), T*(17). *Mechanics: From Newton's Laws to Deterministic Chaos.* Florian Scheck. Springer-Verlag, 1990, xiv + 431 pp, $39.50 (P). [ISBN: 0-387-52715-X] English version of the *Second Enlarged Edition* of the German original developed for a course in classical physics. Rigid body mechanics, dynamics of many-body systems, mechanics following Euler, Lagrange, Hamilton and Jacobi, elements of special relativity, and an introduction to continuous systems. Sophisticated exercises. Translation into English of solutions to exercises is in the works. Prerequisites include linear algebra, old-fashioned advanced calculus, and modern analysis. Appendices with "practical" examples and worked-out examples in text are helpful. JDEK

Applications (Physics), S(16-18), P, L. *Instantons and Four-Manifolds, Second Edition.* Daniel S. Freed, Karen K. Uhlenbeck. Math. Sci. Res. Inst., V. 1. Springer-Verlag, 1991, xxi + 194 pp, $29.95. [ISBN: 0-387-97377-X] Material from the *First Edition* was not substantially revised. However, this newly typeset edition does contain a brief survey of recent advances and an expanded bibliography. (*First Edition,* TR, March 1985.) MU

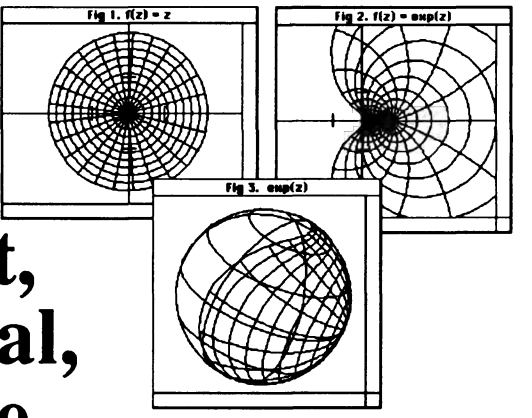
Applications (Physics), S(18), P. *Vortex Methods and Vortex Motion.* Eds: Karl E. Gustafson, James A. Sethian. SIAM, 1991, viii + 212 pp, (P). [ISBN: 0-89871-258-0] Largely discursive; begins with theoretical overview, but devotes most attention to well-illustrated examples. MU

Applications (Physics), P. *Quantum Theory without Reduction.* Eds: Marcello Cini, Jean-Marc Lévy-Leblond. Adam Hilger, 1990, ix + 167 pp, $39. [ISBN: 0-7503-0031-0] One of the conundrums of quantum mechanics is the "reduction of the state vector"—the collapse to a single eigenvector when a measurement is made. Can the theory do without it? Eleven papers from a 1989 conference consider the question. BC

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corsatt, St. Olaf; AD: Amy Davidow, Macalester; BH: Bruce Hanson, St. Olaf; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; JS: John Schue, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

Software and video tapes for the student, professional, and anyone who loves mathematics.



Ask for our latest catalog.

Lascaux Graphics 3220 Steuben Ave.
Bronx, NY 10467 USA (800) 338-0993

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.50 MAA Member: $5.00

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

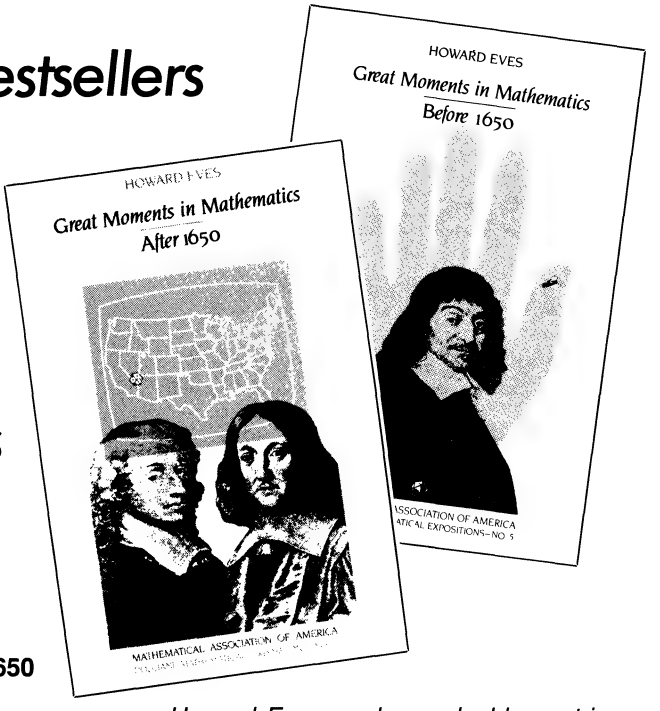
In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

The MAA Bestsellers on the History of Mathematics



GREAT MOMENTS IN MATHEMATICS BEFORE 1650

Howard Eves

This outstanding volume (along with its companion Volume 7) has become an MAA bestseller. It contains 23 fascinating mathematical episodes from antiquity to about 1650.

270 pp., Paper, 1982, ISBN 0-88385-310-8

List: $17.00 MAA Member: $13.00

Catalog Number DEP-05

GREAT MOMENTS IN MATHEMATICS AFTER 1650

Howard Eves

This is a companion volume to Volume 5.

270 pp., Paper, 1982, ISBN 0-88385-311-6

Paper List: $17.00 MAA Member: $13.00
Catalog Number DEP-07 (Paper)

Hardbound List : $27.00
MAA Member: $20.00
Catalog Number DOL-07 (Hardbound)

Both books available in paperbound editions as a package: List: $31.50 MAA Member: $23.00

Catalog Number DEP-057

Howard Eves made a valuable contribution to the Dolciani Mathematical Exposition series . . . The twenty lectures included are a delight to read. They place each "great moment" in its historical context and lay special emphasis on human aspects of each achievement. No algebra or geometry teacher should be without this book.

Tom Walsh in *The Mathematics Teacher*

He has the knack of bringing ancient mathematics alive, at the same time relating it to modern ideas and developments. His work is readable and informative and indeed bears reading a second time. His exposition is clear and he is not content with generalities, but gives actual examples of early methods.

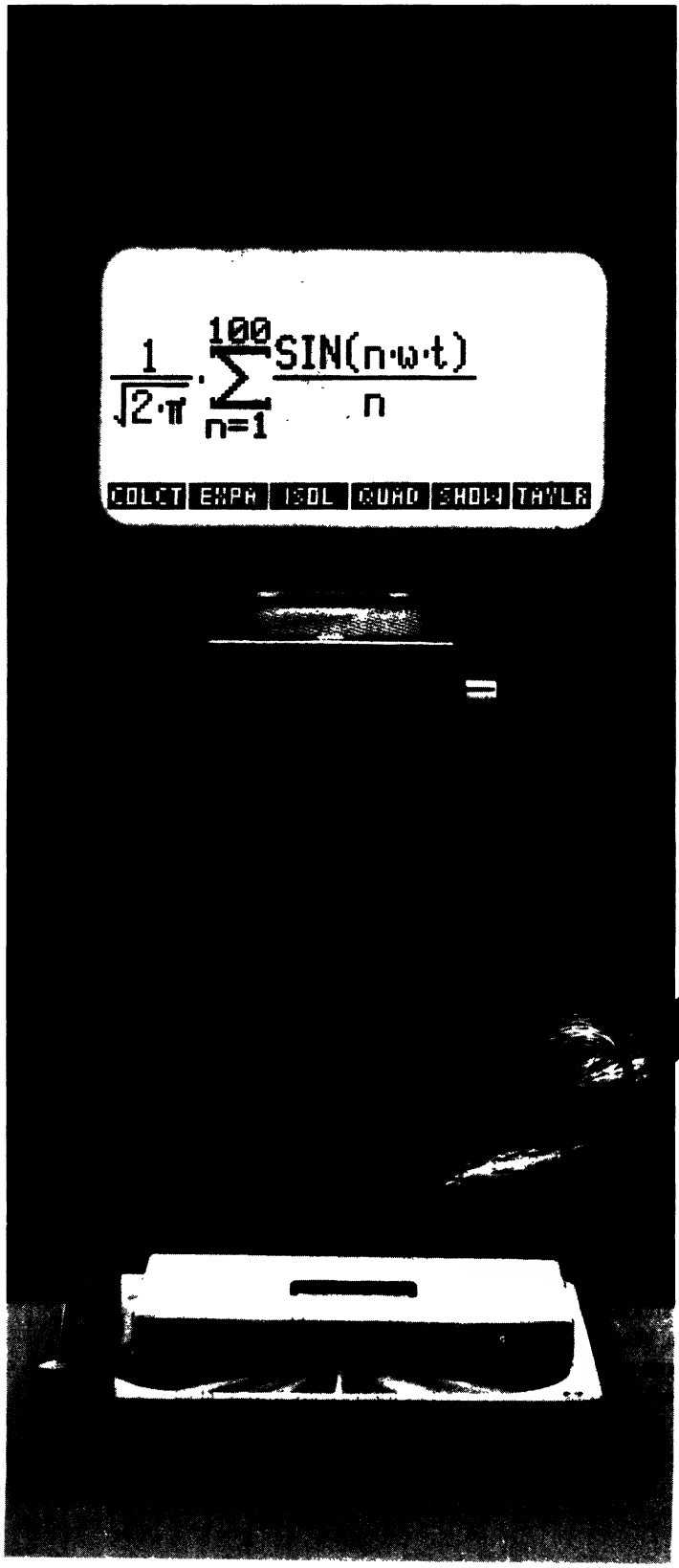
E.H. Lockwood in
The Mathematical Gazette

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Help your students discover more meaningful relationships.



Renewed for '91: a free classroom display device with purchase of 30 calculators.

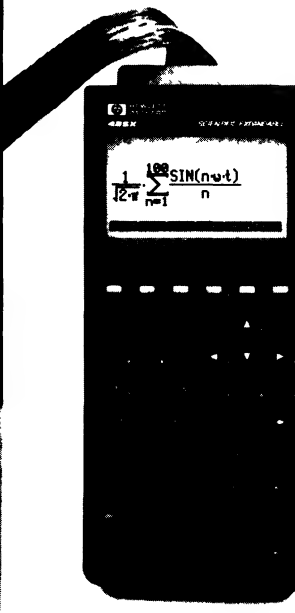
Showing is much more powerful than telling. So we've developed special classroom displays for our most advanced calculators.

The HP 48SX scientific expandable calculator, the cost-effective HP 28S and the new HP 48S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculators.

If your department or students purchase 30 HP 48SX, HP 28S or HP 48S calculators (or a mix of all three), we'll give you free an HP 48SX and plug-in classroom display (a $900 retail value). Or a classroom display version of the HP 28S (a $600 retail value). And free teaching materials.

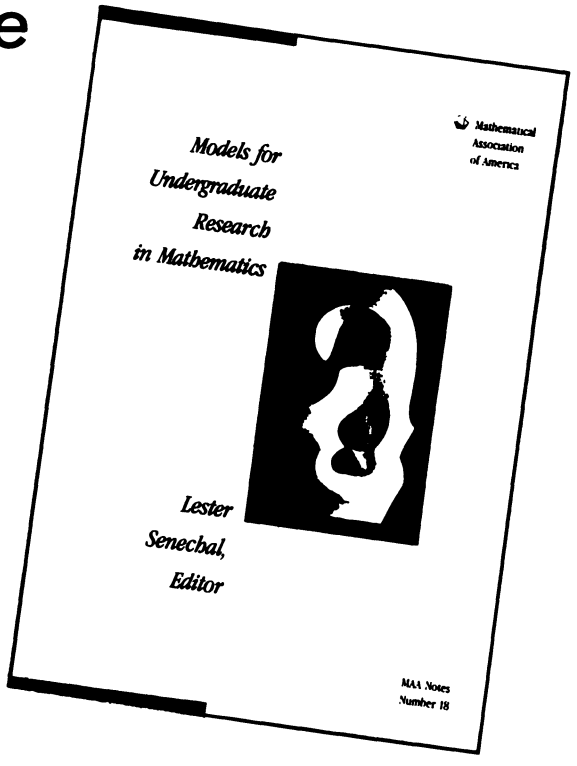
Call **(503) 757-2004** from 8am to 3pm PDT for details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1991, and applies only to college and high school instructors.

HEWLETT PACKARD



© 1991 Hewlett-Packard Company PG12005

Undergraduate Research



MODELS FOR UNDERGRADUATE RESEARCH IN MATHEMATICS

Lester J. Senechal, Editor

While there seems to be no easy way of turning undergraduates into research mathematicians, some schools have run successful programs which increased the research participation of their students. Many such model programs are presented in this book. You will learn how dramatic results were achieved in various intensive summer programs, when full attention was devoted to the research process. You will also learn about successful, well-conceived, academic-year research-oriented programs for undergraduate students. In addition to detailed descriptions of undergraduate research programs by faculty, there are accounts by students of their experiences and also a few sample undergraduate research papers. The final section of the book is a draft of a position paper by the Committee on Science Policy of the Council of Undergraduate Research. It offers some useful guidelines for faculty and administators and implicitly sets some challenges. We have much work ahead of us it its tenets are to become accepted and implemented in the mathematics community.

Faculty in undergraduate colleges are typcially removed from research activity; and undergraduates even in prestigous research universities, may have little opportunity to experience mathematics as a creative process. The cost to mathematics comes when some of our most mathematically gifted students elect careers outside of mathematical science in which they perceive that they will play a more active role.

Order a copy of this important new book and share it with your colleagues. No mathematics department should be without it.

208 pp., Paperbound, 1990
ISBN 0-88385-070-2

List $15.00

Catalog Number NTE-18

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

NEW! VERSION 2

Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE May 29, 1990 Derive, Version 1.6

DERIVE® – A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

NEW MEMBERS JOIN NOW NEW MEMBERS JOIN NO\

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036-1385

Name _____

Mailing Address _____

_____ Zip _____

Employer/School _____

Position (Rank) _____

Employer's City/State _____

Highest Earned Degree _____ Year Degree Earned _____

Institution Awarding Degree _____

Month/Year of Birth _____

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Find the column for your desired combination of MAA journals in the table below. All members receive FOCUS. **Select** the row appropriate for your initial membership period (1 year or 1.5 years) and your status (student or regular). **Circle** your dues in the table. **Write** the amount on the form below.

Journal codes used in the columns of the dues table are: THE AMERICAN MATHEMATICAL MONTHLY = **M**
MATHEMATICS MAGAZINE = **G** THE COLLEGE MATHEMATICS JOURNAL = **J**

Student Membership	**M**	**G**	**J**	**M + G**	**M + J**	**G + J**	**M + G + J**
1 year (Jan.–Dec. 1991)	$32.00	$27.00	$29.00	$ 37.00	$ 39.00	$ 33.00	$ 44.00
1.5 years (July 1991–Dec. 1992)	$50.00	$41.50	$44.50	$ 58.50	$ 60.50	$ 51.50	$ 69.00
Regular Membership							
1 year (Jan.–Dec. 1991)	$63.00	$47.00	$51.00	$ 79.00	$ 83.00	$ 67.00	$ 99.00
1.5 years (July 1991–Dec. 1992)	$99.50	$78.50	$83.50	$121.50	$126.50	$105.50	$148.50

These are **specially discounted rates for new members** provided to help the **MAA** reach a wider audience. They are not available to those who have been members for two years. **Student membership** is available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Student rates apply to unemployed persons who are seeking employment. **Annual dues include subscription prices as follows:** Regular Member $34 **M**, $21 **J**, $17 **G**, $6 **Focus**. Student member $11 **M**, $8 **J**, $6 **G**, $3 **Focus**. These rates are guaranteed for the indicated periods only.

Payment Enclosed $ _____ *U.S. Funds Only* (see below)

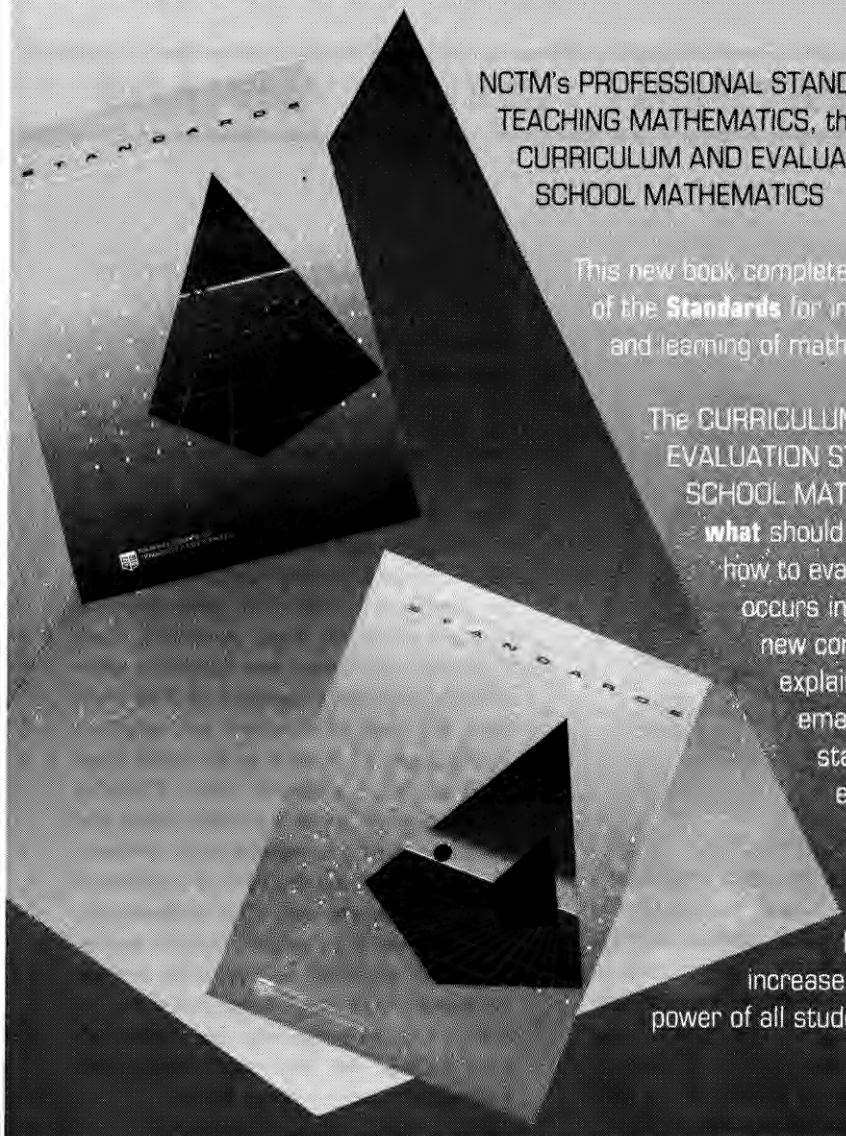
METHOD OF PAYMENT ☐ Check Enclosed ☐ VISA ☐ MasterCard

Card Number _____ Expiration Date |_|_|_|_| *(month/year)*

Interbank Number |_|_|_|_| *(MasterCard only—located above name on card)*

SIGNATURE _____

Just published!



NCTM's PROFESSIONAL STANDARDS FOR TEACHING MATHEMATICS, the companion to the CURRICULUM AND EVALUATION STANDARDS FOR SCHOOL MATHEMATICS

This new book completes the two-document set of the **Standards** for improving the teaching and learning of mathematics in the 1990s.

The CURRICULUM AND EVALUATION STANDARDS FOR SCHOOL MATHEMATICS outlined **what** should be taught and showed how to evaluate the learning that occurs in the classroom. This new companion document explains how to teach mathematics, and it establishes standards for teacher evaluation and professional development.

Used together, these books will help teachers increase the mathematical power of all students.

Special price for the two-book set, **$42.50**
ISBN 0-87353-308-9, #48156

Professional Standards for Teaching Mathematics
200 pp., ISBN 0-87353-307-0, #43956, $25*

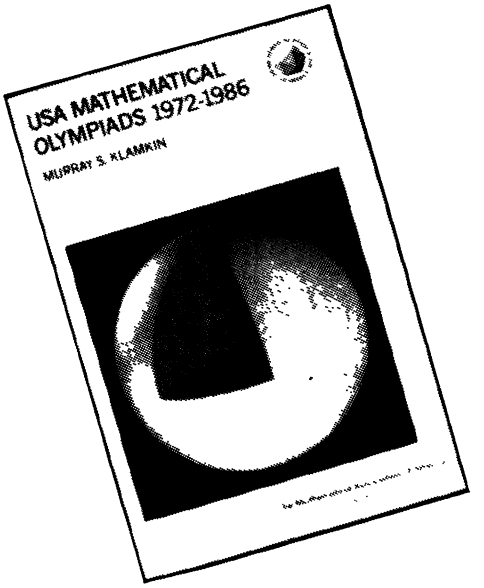
Curriculum and Evaluation Standards for School Mathematics
258 pp., ISBN 0-87353-273-2, #39656, $25
($20 for NCTM individual members)

(Quantity discounts for each book are also available.)

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS
1906 Association Drive, Reston, VA 22091
Tel. (703) 620-9840 or fax (703) 476-2970
To order, call (800) 235-7566

** $10 **for all new** NCTM individual members and **for all** NCTM institutional members until 1 August 1991.*

USA MATHEMATICAL OLYMPIADS



Every year 100 of the most mathematically talented high school students in the country compete in the USA Mathematical Olympiad (USAMO). The USAMO is the third stage of a three-tiered mathematical competition for high school students in the United States and Canada that begins with the AHSME taken by over 400,000 students, continues with the American Invitational Mathematics Exam involving 2,000 students, and culminates in the 100-contestant USAMO.

USA MATHEMATICAL OLYMPIADS 1972-1986, PROBLEMS AND SOLUTIONS

Compiled by Murray S. Klamkin

People delight in working on problems "because they are there, " for the sheer pleasure of meeting a challenge. This is a book full of such delights. In it, Murray S. Klamkin brings together 75 original USA Mathematical Olympiad (USAMO) problems for years 1972-1986, with many improvements, extensions, finger exercises, open problems, references and solutions, often showing alternative approaches. The problems are coded by subject and solutions are arranged by subject as an aid to those interested in a particular field. Contains a glossary of frequently used terms and theorems, and a comprehensive bibliography with items numbered and referred to in brackets in the text. The problems are intriguing and the solutions elegant and informative. Students and teachers will enjoy working these challenging problems. Indeed all those those who are mathematically inclined will find many delights and pleasant challenges in this book.

180 pp., 1988, ISBN-0-88385-634-4

List: $13.50 MAA Member: $12.50

Catalog Number NML-33

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Cambridge University Press

Probability with Martingales
David Williams
1991 272 pp. 40455-X Hardcover $59.50

Integral Equations
A Practical Treatment, from Spectral Theory to Applications
David Porter *and* ***David S.G. Stirling***
Cambridge Texts in Applied Mathematics 5
1990 384 pp. 33151-X Hardcover $69.50
33742-9 Paper $27.95

Integral Equations and Applications
C. Corduneanu
1991 376 pp. 34050-0 Hardcover $89.50

Some Applications of Modular Forms
Peter Sarnak
Cambridge Tracts in Mathematics 99
1990 121 pp. 40245-X Hardcover $29.95

Infinite-Dimensional Lie Algebras
Third Edition
Victor G. Kac
1990 421 pp. 37215-1 Hardcover $69 50

Nonlinear Waves, Solitons and Chaos
E. Infeld *and* ***G. Rowlands***
1990 435 pp. 32111-5 Hardcover $85.00
37937-7 Paper $29.95

Linear Algebra
An Introduction with Concurrent Examples
A.G. Hamilton
1990 328 pp. 32517-X Hardcover $69.50
31042-3 Paper $22.95

The Theory of Distributions
A Nontechnical Introduction
Ian Richards *and* ***Heekyung Youn***
1990 156 pp. 37149-X Hardcover $39.50

Geometry of Low-Dimensional Manifolds
Volumes 1 and 2
Edited by ***S.K. Donaldson*** *and* ***C.B. Thomas***
London Mathematical Society Lecture Note Series 151, 152
Vol. 1: 1991 272 pp. 39978-5 Paper $29.95
Vol 2: 1991 350 pp. 40001-5 Paper $29.95

Steps in Commutative Algebra
R.Y. Sharp
London Mathematical Society Student Texts 19
1990 320 pp. 39338-8 Hardcover $49.50
39732-4 Paper $17.95

A Tribute to Paul Erdős
Edited by ***A. Baker, B. Bollobas,*** *and* ***A. Hajnal***
1991 493 pp. 38101-0 Hardcover $99.50

Transcendental Number Theory
Alan Baker
Cambridge Mathematical Library
1990 175 pp. 39791-X Paper $17.95

Algebraic Curves Over Finite Fields
Error-Correcting Codes and Exponential Sums
Edited by ***Carlos Moreno***
Cambridge Tracts in Mathematics 97
1991 255 pp. 34252-X Hardcover $49.50

Yet Another Introduction to Analysis
Victor Bryant
1990 298 pp. 38166-5 Hardocver $60.00
38835-X Paper $19.95

Linear Analysis
B. Bollobas
1990 220 pp. 38301-3 Hardcover $59.50
38729-9 Paper $19.95

Now in paperback.

Heat Kernels and Spectral Theory
E.B. Davies
Cambridge Tracts in Mathematics
1991 208 pp. 40997-7 Paper $24.95

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.
MasterCard/VISA accepted Prices subject to change

Two Books On Mathematical Ingenuity By Ross Honsberger

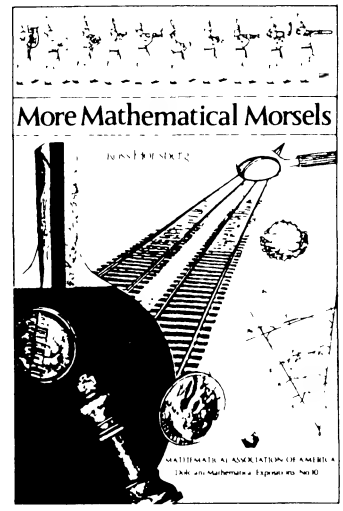
Ross Honsberger is the author of seven books in the Dolciani Mathematical Exposition series, each of which presents problems from algebra, arithmetic, number theory, probability, and geometry, and provides ingenious solutions and/or intriguing results. His most recent addition to the series is MORE MATHEMATICAL MORSELS which is a continuation of the earlier Mathematical Morsels volume published in 1979. Now is your opportunity to own both of these excellent books. The problems presented are meant to be enjoyed, rather than instruct, although instruction is almost always the automatic by-product.

Honsberger says in the Preface of his first volume of morsels, "Mathematics abounds in bright ideas. No matter how long and hard one pursues her, mathematics never seems to run out of exciting surprises. And by no means are these gems to be found only in difficult work at an advanced level. All kinds of simple notions are full of ingenuity. The present volume discusses scores of elementary problems and contains dozens of marvellous ideas."

All of the problems are accessible to anyone with a knowledge of freshman mathematics. Buy both of these excellent books today, and share them with your students.

This is an excellent text...It exposes gifted undergraduate students to a set of elementary problems...whose proofs are ingenious, usually containing beautiful ideas...Its lucid expository style and stimulating mathematical content make its reading a rewarding experience for mathematicians at all levels. (about Mathematical Morsels)

M.S. Cheema in *Mathematical Reviews*



...even the more advanced theoretician will find many of the problems fascinating...

D.T. Swift-Hook in *Physics Education (about Mathematical Morsels)*

NEW NEW NEW NEW NEW NEW

MORE MATHEMATICAL MORSELS

Ross Honsberger

315 pp., Paperbound, 1990,

List: $16.00 MAA Member: $13.50

Catalog Number DOL-10

AN MAA BESTSELLER

MATHEMATICAL MORSELS

Ross Honsberger

249 pp., Hardbound, 1979
ISBN 0-88385-303-5

List: $ 28.00 MAA Member: $ 21.00

Catalog Number DOL-03

Both volumes purchased now

List: $40.00 MAA Member: $29.00

Catalog Number DOL-310

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 6 June–July 1991

Contents

(ISSN 0002-9890)

ARTICLES

Magic Squares and Linear Algebra CHRISTOPHER J. HENRICH 481

An Analytical Description of Some Simple Cases of Chaotic Behaviour JAMES V. WHITTAKER 489

Multiple Pythagorean Number Triples ALBERT FÄSSLER 505

A Very Elementary Proof of the Malgrange-Ehrenpreis Theorem JEAN-PIERRE ROSAY 518

NOTES

On the Zeros of $\tan^{(k)} z$ LI-CHIEN SHEN 524

Two Equal Sums of Three Squares with Equal Products JOHN B. KELLY 527

A Coloring Problem IRA M. GESSEL 530

Norm Inequalities for the Powers of a Matrix MAN KAM KWONG AND A. ZETTL 533

THE TEACHING OF MATHEMATICS

Numerical Linear Algebra on the HP-28 or How to Lie with Supercalculators YVES NIEVERGELT 539

Gram-Schmidt Orthogonalization by Gauss Elimination LYLE PURSELL AND S. Y. TRIMBLE 544

On Finite Groups and Finite Fields J. D. REID 549

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 552

Advanced Problems and Solutions 559

REVIEWS

Mathematics and Its History. By John Stillwell JEFFREY NUNEMACHER 569

Telegraphic Reviews 575

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $32 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

Magic Squares and Linear Algebra

CHRISTOPHER J. HENRICH, *Concurrent Computer Corporation, Tinton Falls, NJ 07724*

CHRISTOPHER J. HENRICH: After receiving my doctorate at Harvard in 1968, I taught at Columbia and at Fordham. Since 1975, I have been working for Concurrent Computer Corporation (which was formerly Interdata, and until 1985 a subsidiary of the Perkin-Elmer Corporation). My work has been concentrated in compilers and mathematical software; I am now a member of the Ada compiler development team. My research interests have included partial differential equations and the representation theory of compact Lie groups.



1. Introduction. A *magic square of order n* is an arrangement of the numbers $\{1, \ldots, n^2\}$ in an $n \times n$ square, so that each row, column, or diagonal adds up to the same sum S. Because $nS = 1 + \cdots + n^2$, we have

$$S = n(n^2 + 1)/2. \tag{1.1}$$

Magic squares are a perennial topic of recreational mathematics. They tantalize us with a suggestion of hidden pattern. With a few concepts from linear algebra, we can find a surprisingly simple structure that underlies many magic squares, including some famous examples. We shall use finite vector spaces; in this paper, we only use finite-dimensional vector spaces over the field $\mathbf{F}_2$ with two elements. Familiarity with the first chapter of [3] should be sufficient to enable the student to understand what an affine magic square is and how to construct one.

We shall start by examining the square of order 4 that appears in Dürer's engraving *Melancolia*. (See, for instance, [1, p. 147].) This will lead us to the description of the family of *affine magic squares* of order 4. All the magic squares of order 4 are known, and we shall see what portion of them are affine. We briefly discuss two squares found by Benjamin Franklin, to show how the theory extends beyond squares of order 4.

2. A close look at Dürer's square. Consider a four by four square. Number the rows and columns, starting with 0 and using base-2 notation. The row and column numbers then label the positions in the square uniquely; a label contains four digits, which may be 0 or 1. (See FIGURE 1.)

Now let us forget that these 0's and 1's came from positional notation for integers, and regard them as values in $\mathbf{F}_2$. Thus we have identified the cells of the square with the elements of $\mathbf{F}_2^4$.

	00	01	10	11
00	0000	0001	0010	0011
01	0100	0101	0110	0111
10	1000	1001	1010	1011
11	1100	1101	1110	1111

FIG. 1

We shall be interested in subspaces of $\mathbf{F}_2^4$. A notation for the vector subspace spanned by several vectors will be that list of vectors, enclosed in angle brackets. We will use the common notation $\{e_1, e_2, e_3, e_4\}$ for the natural basis of $\mathbf{F}_2^4$. Our motive for looking at subspaces is that they correspond to features of interest in the study of magic squares. For instance, the first row is $\langle e_3, e_4 \rangle$, and the affine subspaces parallel to this correspond to the other rows. Similarly, the columns of the square are of the form $x + \langle e_1, e_2 \rangle$ for suitable choices of x in $\mathbf{F}_2^4$. The main diagonal is $\langle e_1 + e_3, e_2 + e_4 \rangle$; the affine spaces parallel to this are the other diagonal and two broken diagonals, represented in FIGURE 2. Some magic squares, called *quasi-pandiagonal*, are magic for the broken diagonals of FIGURE 2 but not for those of FIGURE 3. The latter are another parallel family of affine spaces. The reader should look at some of the other two-dimensional subspaces of $\mathbf{F}_2^4$, and their parallel families of affine subspaces.

Now let us consider Dürer's square; it is transcribed here in FIGURE 4.

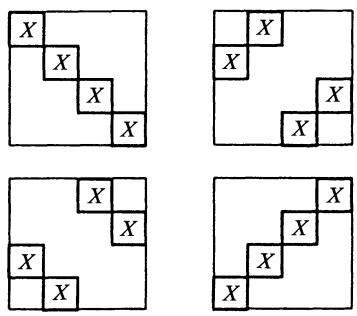


FIG. 2

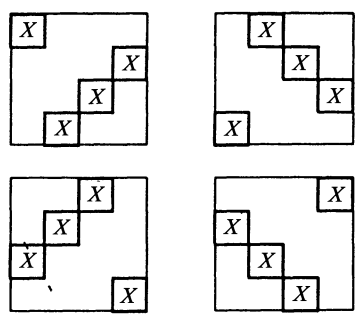


FIG. 3

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

FIG. 4

0	1	0	1
1	0	1	0
1	0	1	0
0	1	0	1

FIG. 5

Some of the four-dimensional order hidden in this square comes to light if we look at the parity of the values in its cells, in FIGURE 5.

This, in a humble way, also has the magic property: in every row, column, and diagonal the sum is 2. It is also related to the vector space structure of $\mathbf{F}_2^4$. In fact, if the 0's and 1's are understood as belonging to $\mathbf{F}_2$, then FIGURE 5 is a linear function from $\mathbf{F}_2^4$ to $\mathbf{F}_2$. We need some more notation; we shall let $\{f_1, f_2, f_3, f_4\}$ be the dual basis to $\{e_1, e_2, e_3, e_4\}$, so that $f_i(e_j)$ is 1 if $i = j$ and 0 otherwise. Then FIGURE 5 displays

$$f_1 + f_2 + f_4.$$

Because this is a linear function, on any vector subspace of $\mathbf{F}_2^4$ it either is constant (equal to 0) or takes the values 0 and 1 equally often. If the latter is true on a vector subspace, then it is equally true on any parallel affine subspace. This function is non-constant on the subspaces corresponding to lines, columns, and diagonals of the square; therefore on each of these affine spaces it takes the values 0 and 1 equally often. Here is a general explanation of why FIGURE 5 has the magic property. It is worth restating in slightly greater generality:

PROPOSITION 1. *Let ψ be a linear function on $\mathbf{F}_2^4$ with values in $\mathbf{F}_2$; let $\mathbf{E}$ be a linear subspace of $\mathbf{F}_2^4$ which is not contained in the null-space of ψ. Then, on any affine subspace parallel to $\mathbf{E}$, the function ψ takes the values 0 and 1 equally often.*

Can similar reasoning account for all the magic in Dürer's square? We can look beyond the parity of the numbers in the square by representing them in base 2. If we try this, we find that it is tidier to subtract 1 from each number first. Then they range from 0 to 15, and their base-2 representations are 0000 through 1111. FIGURE 6 shows the result.

1111	0010	0001	1100
0100	1001	1010	0111
1000	0101	0110	1011
0011	1110	1101	0000

FIG. 6

The respective binary digits in each cell are easier to understand if they are written in separate squares, as in FIGURE 7, and regarded as taking values in $\mathbf{F}_2$. (We shall trust the reader to determine from context when a 0 or 1 is intended to belong in $\mathbf{F}_2$ and when it belongs in $\mathbf{Z}$.) Each digit position determines a function of positions in the square. They are labelled V_j, $j = 1, \ldots, 4$, so that the correspond-

$$V_1: \begin{matrix}1&0&0&1\\0&1&1&0\\1&0&0&1\\0&1&1&0\end{matrix} \qquad V_2: \begin{matrix}1&0&0&1\\1&0&0&1\\0&1&1&0\\0&1&1&0\end{matrix} \qquad V_3: \begin{matrix}1&1&0&0\\0&0&1&1\\0&0&1&1\\1&1&0&0\end{matrix} \qquad V_4: \begin{matrix}1&0&1&0\\0&1&0&1\\0&1&0&1\\1&0&1&0\end{matrix}$$

FIG. 7

ing number in FIGURE 4 is

$$W(x) = 8V_1(x) + 4V_2(x) + 2V_3(x) + V_4(x) + 1. \tag{2.2}$$

The functions in FIGURE 7 turn out to be linear functions, modified by a constant term; that is *affine functions*. Specifically,

$$\begin{aligned} V_1 &= \quad f_2 + f_3 + f_4 + 1 \\ V_2 &= f_1 \quad + f_3 + f_4 + 1 \\ V_3 &= f_1 + f_2 + f_3 \quad + 1 \\ V_4 &= f_1 + f_2 \quad + f_4 + 1 \end{aligned} \tag{2.3}$$

They are given in tabular form in FIGURE 8.

	f_1	f_2	f_3	f_4	1
V_1	0	1	1	1	1
V_2	1	0	1	1	1
V_3	1	1	1	0	1
V_4	1	1	0	1	1

FIG. 8

Now we can see, very clearly, why Dürer's square has to have the magic properties that it does. Each of the value functions is non-constant on the subspaces that correspond to rows, columns, and diagonals. On any affine subspace $\mathbf{E}$ of $\mathbf{F}_2^4$, an affine function W that is not constant must, as we have seen, take the values 0 and 1 equally often. Thus each of the V_n takes the value 1 on exactly two elements in each row, or in each column, or in each diagonal. This fact accounts for the uniformity of the sums of $W(x)$ over rows, columns, and diagonals. The numbers $W(x)$ include all of $\{1,\ldots,16\}$ with no duplications, because the linear parts of the value functions are linearly independent.

It is natural to generalize the properties of Dürer's square. At the same time, we can summarize our discoveries in a series of easy propositions.

DEFINITION 1. *If V_j, $j = 1,\ldots,4$ are affine functions on $\mathbf{F}_2^4$ with values in $\mathbf{F}_2$, then the map W from $\mathbf{F}_2^4$ to $\mathbf{Z}$ given by equation* (2.2) *is an affine square.*

In discussion of a particular affine square W, we shall denote the linear part of V_j by ψ_j.

DEFINITION 2. *Let $V_1,\ldots,V_4$ be as in definition* 1, *and let W be the affine square determined by them; let $\mathbf{E}$ be a linear subspace of $\mathbf{F}_2^4$. Then $\mathbf{E}$ is magic for W if each of the linear functions ψ_j is non-zero on $\mathbf{E}$.*

PROPOSITION 2. *Let $V_1, \ldots, V_4$ be as in Definition* 1, *and let* **E** *be a linear subspace of* $\mathbf{F}_2^4$ *which is magic for the affine square W determined by the V_j. Then W has uniform sums on the affine subspaces parallel to* **E**.

Proof. Let **E** have dimension d. By Proposition 1, each of the functions ψ_j takes the values 0 and 1 each 2^{d-1} times on every affine space parallel to **E**. Clearly the same is true of V_j. Therefore the sum (in **Z**) of V_j over every such affine subspace is 2^{d-1}; the conclusion follows at once in view of equation (2.2).

DEFINITION 3. *Let $V_1, \ldots, V_4$ be as in Definition* 1, *and let W be the affine square determined by the V_j. Then W is nonsingular if $\psi_1, \ldots, \psi_4$ are linearly independent.*

PROPOSITION 3. *Let $V_1, \ldots, V_4$ be as in Definition* 1, *and suppose that the affine square W determined by the V_j is non-singular. Then the numbers in W are exactly* $\{1, \ldots, 16\}$ *with no omissions or duplications.*

Proof. Because $1 \leq W(x) \leq 16$ for $x \in \mathbf{F}_2^4$, and there are 16 points in $\mathbf{F}_2^4$, it is sufficient to show that no two of them have the same value of W. Now suppose x and y are members of $\mathbf{F}_2^4$ and $V_j(\mathrm{x}) = V_j(\mathrm{y})$, for $j = 1, \ldots, 4$. Then we have $\psi_j(\mathrm{x}) = \psi_j(\mathrm{y})$ for all j. But by hypothesis the ψ_j are linearly independent, and therefore span $\mathbf{F}_2^{4*}$. It follows that x = y.

DEFINITION 4. *An affine magic square is an affine square which is nonsingular, and for which the subspaces $\langle e_1, e_2 \rangle$, $\langle e_3, e_4 \rangle$, and $\langle e_1 + e_2, e_3 + e_4 \rangle$ are magic.*

PROPOSITION 4. *Every affine magic square is a "pandiagonal" or "quasi-pandiagonal" magic square.*

Proof. This is just a combination of Propositions 2 and 3.

3. How many affine magic squares exist? To answer this question, we must look closely at affine functions on $\mathbf{F}_2^4$. Their linear parts are the most important consideration.

For a linear function to be non-constant on the prescribed subspaces, it must satisfy certain constraints. Because the "row" subspace is magic, it must be nonzero on either e_3 or e_4. Likewise, because the "column" subspace is magic, it must be nonzero on either e_1 or e_2. Finally, for the "diagonal" subspace to be magic, it must be nonzero on either $e_1 + e_3$ or $e_2 + e_4$. Six linear functions satisfy these constraints:

$$L_1 = f_2 + f_3; \quad L_2 = f_1 + f_3 + f_4; \quad L_3 = f_1 + f_2 + f_4;$$
$$L_4 = f_2 + f_3 + f_4; \quad L_5 = f_1 + f_4; \quad L_6 = f_1 + f_2 + f_3.$$

To determine an affine magic square, we must select a set of four functions among which there is no linear relation. Accordingly, we look for linear relations among the six eligible functions. A relation among three functions, $f + f' + f'' = 0$, is equivalent to the equation $f + f' = f''$. There are two such relations:

$$L_1 + L_2 + L_3 = 0;$$
$$L_4 + L_5 + L_6 = 0.$$

A relation among four functions can be expressed as $f + f' = g + g'$. However, one finds that a nonzero function that is not eligible can be expressed as a sum of two eligible functions in only one way. Therefore there are no four-term relations among the eligible functions.

From six eligible functions, a set of four can be selected in 15 ways. Each of the three-term relations excludes 3 of these sets. Thus there remain 9 linearly independent sets of eligible linear functions. Now a set of linear functions determines 16 sets of four affine functions, because each linear function can be given a constant term of 0 or 1. And each such set can be arranged in $4! = 24$ ways. Thus the number of affine magic squares of order 4 is $9 \cdot 16 \cdot 24 = 3456$. If we take the 8 symmetries of the geometrical square into account, the number of *distinct* affine magic squares is $3456/8 = 432$.

A complete enumeration of the magic squares of order 4 has been known since the seventeenth century, when it was found by Bernard Frénicle de Bessey. Bensen and Jacoby [2] checked his work using a computer. A further detailed study, with analysis and proofs, is found in [4]. In this census, two squares are identified if they differ only by one of the 8 symmetries of a geometrical square. There are 880 distinct magic squares; 48 are pandiagonal, and another 384 are quasi-pandiagonal. Thus we have a converse to Proposition 4.

PROPOSITION 5. *Every "pandiagonal" or "quasi-pandiagonal" magic square of order* 4 *is an affine magic square.*

4. Benjamin Franklin's examples. The theory and notation of section 2 are easily extended to describe affine magic squares whose order is any power of 2. Examples of such squares have long been known. Benjamin Franklin constructed magic squares of order 8 and 16. An account of them is given in [**1**, pp. 89–112]. The square of order 8 is transcribed as FIGURE 9. If we subtract 1 from each number and convert to base 2, we get FIGURE 10, which is analogous to FIGURE 6.

52	61	4	13	20	29	36	45
14	3	62	51	46	35	30	19
53	60	5	12	21	28	37	44
11	6	59	54	43	38	27	22
55	58	7	10	23	26	39	42
9	8	57	56	41	40	25	24
50	63	2	15	18	31	34	47
16	1	64	49	48	33	32	17

FIG. 9

	000	001	010	011	100	101	110	111
000	110011	111100	000011	001100	010011	011100	100011	101100
001	001101	000010	111101	110010	101101	100010	011101	010010
010	110100	111011	000100	001011	010100	011011	100100	101011
011	001010	000101	111010	110101	101010	100101	011010	010101
100	110110	111001	000110	001001	010110	011001	100110	101001
101	001000	000111	111000	110111	101000	100111	011000	010111
110	110001	111110	000001	001110	010001	011110	100001	101110
111	001111	000000	111111	110000	101111	100000	011111	010000

FIG. 10

	f_1	f_2	f_3	f_4	f_5	f_6	1
V_1	0	0	1	1	1	0	1
V_2	0	0	1	0	1	0	1
V_3	0	0	1	0	0	1	0
V_4	1	1	1	0	0	1	0
V_5	0	1	1	0	0	1	1
V_6	1	1	0	0	0	1	1

FIG. 11

We can check that the square is affine; the functions which determine it are tabulated in FIGURE 11. Several subspaces of dimension 2 and 3 in $\mathbf{F}_2^6$ are magic for this square. Among the 2-dimensional such subspaces are these:

$$\langle e_3, e_6\rangle;$$
$$\langle e_2 + e_3, e_6\rangle;$$
$$\langle e_3, e_5 + e_6\rangle;$$
$$\langle e_2 + e_3, e_5 + e_6\rangle.$$

One can check that every contiguous 2-by-2 block in the square corresponds to an affine space parallel to one of these subspaces.

Franklin observed that every bent row of 8 numbers ascending and descending diagonally totals 260, for example the row from 52 descending to 54 and from 43 ascending to 45. This is the image of the vector subspace

$$\langle e_3 + e_6, e_2 + e_5, e_2 + e_3 + e_4\rangle,$$

which is magic for this square. This fact accounts for only 4 of the 10 bent rows. What explains the magic of the row from 14 descending to 10, and from 23

200	217	232	249	8	25	40	57	72	89	104	121	136	153	168	185
58	39	26	7	250	31	218	199	186	167	154	135	122	103	90	71
198	219	230	251	6	27	38	59	70	91	102	123	134	155	166	187
60	37	28	5	252	229	220	197	188	165	156	133	124	101	92	69
201	216	233	248	9	24	41	56	73	88	105	120	137	152	169	184
55	42	23	10	247	234	215	202	183	170	151	138	119	106	87	74
203	214	235	246	11	22	43	54	75	86	107	118	139	150	171	182
53	44	21	12	245	236	213	204	181	172	149	140	117	108	85	76
205	212	237	244	13	20	45	52	77	84	109	116	141	148	173	180
51	46	19	14	243	238	211	206	179	174	147	142	115	110	83	78
207	210	239	242	15	18	47	50	79	82	111	114	143	146	175	178
49	48	17	16	241	240	209	208	177	176	145	144	113	112	81	80
196	221	228	253	4	29	36	61	68	93	100	125	132	157	164	189
62	35	30	3	254	227	222	195	190	163	158	131	126	99	94	67
194	223	226	255	2	31	34	63	66	95	98	127	130	159	162	191
64	33	32	1	256	225	224	193	192	161	160	129	128	97	96	65

FIG. 12

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	1
V_1	0	0	0	1	1	1	0	0	1
V_2	0	0	0	1	0	1	0	0	1
V_3	0	0	0	1	0	0	1	0	0
V_4	0	0	0	1	0	0	0	1	0
V_5	1	1	0	1	0	0	0	1	0
V_6	0	1	0	1	0	0	0	1	1
V_7	1	1	1	1	0	0	0	1	1
V_8	1	1	0	0	0	0	0	1	1

FIG. 13

ascending to 19? This is not an affine subspace; but it is the union of two affine subspaces. Indeed, the cells 14 + 60 + 37 + 19 lie in an affine subspace parallel to

$$\langle e_2 + e_3 + e_6, e_4 + e_5 + e_6 \rangle;$$

and the cells 59 + 10 + 23 + 38 lie in an affine space parallel to

$$\langle e_1 + e_2 + e_3 + e_6, e_4 + e_5 + e_6 \rangle.$$

Both subspaces are magic for Franklin's square.

Franklin's square of order 16 is reproduced in FIGURE 12, and the affine functions which generate it are tabulated in FIGURE 13. Franklin was justly proud of this square, calling it "the most magically magical of any magic square ever made by any magician." [1 p. 93] If you explore the affine subspaces of $\mathbf{F}_2^8$ which are magic for this square, you may agree that he was not exaggerating.

REFERENCES

1. W. S. Andrews, Magic squares and cubes, Dover, New York, 1960.
2. William H. Benson and Oswald Jacoby, New Recreations With Magic Squares, Dover, New York, 1976.
3. P. Halmos, Finite-dimensional Vector Spaces, Springer-Verlag, New York, 1974.
4. Kathleen Ollerenshaw and Hermann Bondi, Magic squares of order 4, *Phil. Trans. Royal Soc. London Ser. A*, 306 (1982) 443–532.

An Analytical Description of Some Simple Cases of Chaotic Behaviour

JAMES V. WHITTAKER
Department of Mathematics, University of British Columbia, Vancouver V6T 1W5 Canada

1. Introduction. Considerable attention has been paid to the sets of points z_0 in the complex plane such that the n-th iterate $f^n(z_0)$ of the function

$$f(z) = z^2 - \mu$$

evaluated at the point $z = z_0$ should remain bounded as n tends to ∞. We have only to mention the book [**14**] by Mandelbrot which has rendered this subject accessible to a much wider audience with its fine illustrations and graphical descriptions. A recent wave of activity in this area has generated, among others, the papers by Barnsley et al. [**1**, **2**, **3**, **4**] and by Mandelbrot [**15**, **16**]. A problem which is closely related to the one just described is to analyze the behaviour of the n-th iterate $F^n(x_0)$ of the function

$$F(x) = Mx(1 - x)$$

evaluated at the point x_0 in the interval $0 \leq x \leq 1$. The survey articles by Blanchard [**5**] and Whitley [**18**] and the book by Preston [**17**] provide a comprehensive overview of the problem for this, as well as for other more general classes of functions. The behaviour exhibited by the n-th iterates of f and F as n tends to ∞ in both of these situations has received the general appellation 'chaos'.

In the special cases $\mu = 2$ and $M = 4$ we shall derive explicit analytical formulas for the n-th iterates of the functions f and F which have received some attention in the book by Devaney [**7**, pp. 39–53]. As a result, we shall be able to describe exactly the sets of points z_0 such that $f^n(z_0)$ remains bounded as n tends to ∞ on the interval $-2 \leq z \leq 2$, a fact that was already pointed out by Julia in his original paper [**11**, p. 186]. We shall then turn our attention to the orbits of the function F in the interval $0 \leq x \leq 1$, that is to say, the sets of points $F^n(x_0)$ as n runs over the positive integers. We shall show that the set of points x_0 is dense in $[0, 1]$ in each of the following cases: the orbit of x_0 is periodic; the orbit of x_0 is eventually absorbed into a fixed point; the orbit of x_0 is dense in $[0, 1]$.

These results are then extended to certain polynomial functions of higher degree which are related to the Chebyshev polynomials, as well as to certain rational functions which are related to the duplication formulas of the circular and Jacobian elliptic functions. In this connection, the work of Lattès [**12**] should also be mentioned. The author is much indebted to Prof. Benoit B. Mandelbrot for helpful comments and references.

2. An analytical formula. We shall take as our point of departure the function

$$X = F(x) = 4x(1 - x)$$

which maps the unit interval $0 \leq x \leq 1$ onto itself. The close connection between this and the function f defined earlier is easily seen from the change of variable defined by

$$z = \phi(x) = 2 - 4x, \qquad Z = \phi(X) = 2 - 4X,$$

for the resulting function is simply

$$Z = f(z) = z^2 - 2.$$

Evidently $Z = \phi(X) = f(z)$ implies that $\phi F = f\phi$, and ϕ maps the interval $[0, 1]$ onto the interval $[-2, 2]$. Furthermore, the function

$$z = h(u) = 2\cos u$$

maps the interval $0 \leq u \leq \pi$ onto $[-2, 2]$. If we now define the function $g(u) = 2u$, then an easy calculation will enable us to verify that

$$fh(u) = (2\cos u)^2 - 2 = 2(2\cos^2 u - 1) = 2\cos 2u = hg(u),$$

that is to say, $fh(z) = hg(z)$ throughout their common domain. A simple iteration of this relation gives us

$$f^2h = ffh = fhg = hgg = hg^2,$$

as well as the general result $f^nh = hg^n$ for every positive integer n. Since $g^n(u) = 2^nu$ for each positive integer n, this latter result can then be stated analytically in the form

$$f^n(2\cos u) = 2\cos g^n(u) = 2\cos(2^nu)$$

for $0 \leq u \leq \pi$. We shall inscribe this result in the following theorem.

THEOREM 1. *If* $f(z) = z^2 - 2$ *for* $-2 \leq z \leq 2$, *then*

$$f^n(2\cos u) = 2\cos(2^nu)$$

for each positive integer n *and* $0 \leq u \leq \pi$.

COROLLARY. *If* $F(x) = 4x(1 - x)$ *for* $0 \leq x \leq 1$, *then*

$$F^n(\sin^2 u) = \sin^2(2^nu)$$

for each positive integer n *and* $0 \leq u \leq \pi/2$.

Proof. Since the relation $\phi F = f\phi$ implies $F^n = \phi^{-1}f^n\phi$, we have only to observe that if we put $x = \sin^2 u$, then we will have $\phi(x) = 2\cos 2u$ and

$$F^n(x) = \phi^{-1}(f^n(2\cos 2u)) = \phi^{-1}(2\cos 2^n2u)) = \frac{1 - \cos(2^n2u)}{2} = \sin^2(2^nu).$$

Since the function $f(z) = z^2 - 2$ is analytic throughout the complex plane, it follows that the analytical formula of Theorem 1 which was verified on the segment $0 \leq u \leq \pi$ must remain valid for all complex values of u according to the principle of the permanence of analytical forms. Moreover, if we put $u = v + iw$ and recall that

$$|\cos u|^2 = \cos^2 v + \sinh^2 w,$$

we then have

$$|\cos(2^nu)|^2 = \cos^2(2^nv) + \sinh^2(2^nw)$$

for each positive integer n, from which it follows that if the imaginary part w of u is different from 0, then $f^n(2\cos u)$ must tend to ∞ as n tends to ∞. Furthermore, the relation

$$\cos u = \cos v \cosh w - i \sin v \sinh w$$

reminds us that the mapping $U = 2\cos u$ sends the line $v = \pi$ to that part of the negative real axis defined by $U \leq -2$, and for such values of U it now follows that $f^n(U)$ will tend to ∞ as n tends to ∞, with the exception of the point $U = -2$. Likewise, the cosine mapping sends the line $v = 0$ to that part of the real axis defined by $U \geq 2$, and for such values of U it follows that $f^n(U)$ will tend to ∞ as n tends to ∞, with the exception of the point $U = 2$. Thus, the only points z_0 in the complex plane such that $f^n(z_0)$ remains bounded as n runs over the positive integers are those in the interval $[-2, 2]$. This phenomenon was verified by Julia in his original paper [**11**, p. 186].

Most of the previous results can be carried over to certain higher degree polynomials $f(z)$. We shall retain the function $z = h(u) = 2\cos u$ but now let $g(u) = ku$ where k is some positive integer. Evidently

$$hg(u) = 2\cos ku = 2T_k(\cos u),$$

where T_k is the Chebyshev polynomial of degree k defined by

$$T_k(z) = \cos(k \arccos z).$$

The leading coefficient of T_k is 2^{k-1}, so if we want a polynomial f_k with leading coefficient 1, then we must put $f_k(z) = 2T_k(z/2)$. From this it readily follows that

$$f_k h(u) = 2T_k(\cos u) = 2\cos ku = hg(u).$$

The rest of the derivation now follows the path mapped out for the case $k = 2$ and leads to the result stated in the next theorem. A few of these polynomials $f_k(z)$ are listed below:

$$z^3 - 3z, \qquad z^4 - 4z^2 + 2, \qquad z^8 - 8z^6 + 20z^4 - 16z^2 + 2.$$

THEOREM 2. *If T_k denotes the k-th Chebyshev polynomial and $f_k(z) = 2T_k(z/2)$, then*

$$f_k^n(2\cos u) = 2\cos(2^n ku)$$

for all positive integers n and all complex values of u.

3. A dynamical model. Many authors such as Collet and Eckmann [**8**] and Whitley [**18**] have taken a different view of the problems described above and regarded the iterates f^n of f as the various phases of a dynamic system. It is apparent from Theorem 1 that we can embed f^n in a continuously evolving system f^t for all real values of the time t according to the formula

$$f^t(2\cos z) = 2\cos(2^t z).$$

This relation can be cast into a more convenient form by means of the change of time scale defined by $t = s/\log 2$, and our system is then governed by the formula

$$f^s(2\cos z) = 2\cos(e^s z).$$

If we put

$$x(s) = 2\cos(ae^s), \qquad y(s) = 2ae^s \sin(ae^s), \qquad z(s) = 2a^2e^{2s}\cos(ae^s), \quad (1)$$

where a is some constant, then repeated differentiation yields the equations

$$x' = -y, \qquad y' = y + z, \qquad z' = 2z - a^2e^{2s}y.$$

If we solve for a^2e^{2s} in the relation $z = a^2e^{2s}x$ and substitute into the expression

for z', then we will arrive at the following autonomous system of nonlinear ordinary differential equations:

$$x' = -y, \qquad y' = y + z, \qquad z' = 2z - \frac{yz}{x}. \tag{2}$$

Of course, the autonomous system (2) can be reduced to the single third-order equation

$$x''' = 3x'' + \frac{x'x''}{x} - \frac{x'^2}{x} - 2x'.$$

Let us now examine an arbitrary trajectory $(x(s), y(s), z(s))$ of the system (2) with initial values

$$x(0) = x_0, \qquad y(0) = y_0, \qquad z(0) = z_0.$$

We shall show that this trajectory lies on a rational surface S and spirals outward on S with increasing time s. Since the right hand sides of the equations in (2) are all continuously differentiable everywhere except when $x = 0$, it follows from the classical existence and uniqueness theorems for ordinary differential equations that the trajectory passing through (x_0, y_0, z_0) must be unique provided that $x_0 \neq 0$. From

$$\begin{aligned}\frac{d}{ds}(xz + y^2) &= x'z + xz' + 2yy' \\ &= -yz + x\left(2z - \frac{yz}{x}\right) + 2y(y + z) \\ &= 2(xz + y^2)\end{aligned}$$

we infer that

$$x(s)z(s) + (y(s))^2 = (x_0z_0 + y_0^2)e^{2s}. \tag{3}$$

Henceforth, we shall assume that $x_0, z_0 > 0$ and observe that

$$\frac{d}{ds}\frac{z}{x} = \frac{xz' - zx'}{x^2} = \frac{x(2z - yz/x) - z(-y)}{x^2} = 2\frac{z}{x},$$

which implies that

$$\frac{z(s)}{x(s)} = \frac{z_0}{x_0}e^{2s}.$$

If we put

$$\mu^2 = \frac{x_0}{z_0}(x_0z_0 + y_0^2)$$

and eliminate e^{2s} in our latest results, we will then find that

$$\frac{z}{x} = \frac{1}{\mu^2}(xz + y^2).$$

This enables us to solve for z with the result that

$$z = \frac{xy^2}{\mu^2 - x^2}. \tag{4}$$

This means that the trajectory must lie on the rational surface (4) in 3-space, and so the trajectory must be nowhere dense in 3-space.

A general analytical solution of the system (2) can easily be obtained by substituting the expressions

$$z = \frac{z_0}{x_0} e^{2s} x, \qquad y = -x'$$

into (3) so as to get the single equation

$$\frac{z_0}{x_0} e^{2s} x^2 + x'^2 = \frac{z_0}{x_0} \mu^2 e^{2s},$$

which simplifies to the form

$$x^2 + \frac{x_0}{z_0} \mu^2 e^{-2s} x'^2 = \mu^2. \tag{5}$$

A simple integration or direct substitution will easily verify that the solution to (5) with initial condition $x(0) = x_0$ is

$$x(s) = \mu \cos\left(\sqrt{\frac{z_0}{x_0}} \frac{e^s - 1}{\mu} + \arccos \frac{x_0}{\mu} \right).$$

The complete solution of (2) now follows from

$$y = -x' = \sqrt{\frac{z_0}{x_0}}\, e^s \sin\left(\sqrt{\frac{z_0}{x_0}} \frac{e^s - 1}{\mu} + \arccos \frac{x_0}{\mu} \right),$$

$$z = y' - y = \frac{z_0}{\mu x_0} e^{2s} \cos\left(\sqrt{\frac{z_0}{x_0}} \frac{e^2 - 1}{\mu} + \arccos \frac{x_0}{\mu} \right).$$

Thus, when s assumes such values that

$$\sqrt{\frac{z_0}{x_0}} \frac{e^s - 1}{\mu} + \arccos \frac{x_0}{\mu} = 2n\pi$$

for some integer n, we will then have

$$x(s) = \mu, \qquad y(s) = 0, \qquad z(s) = \frac{z_0}{\mu x_0} e^{2s}.$$

This means that if S denotes that part of the surface defined by (4) whose domain lies in the infinite strip $x^2 < \mu^2$, then trajectories on S can cross over the x-axis only when they lie either directly above the point $(\mu, 0, 0)$ or else directly below the point $(-\mu, 0, 0)$. Since $y_0 = 0$ implies $z_0 = 0$ in (4), we must exclude all other initial values with $y_0 = 0$ besides the ones just mentioned, so the surface S ought to have an open slit above and below the segment joining the points $(\mu, 0, 0)$ and $(-\mu, 0, 0)$. At the same time we must adjoin the vertical halfline above $(\mu, 0, 0)$ as well as the vertical halfline below $(-\mu, 0, 0)$, so S will remain a connected surface but no longer simply connected.

For any prescribed values of x_0 and μ, we shall choose y_0 and z_0 so that

$$\sqrt{\frac{z_0}{x_0}} = \mu \arccos \frac{x_0}{\mu}, \qquad y_0 = \sqrt{\mu^2 - x_0^2} \arccos \frac{x_0}{\mu}, \tag{6}$$

where x_0, z_0 are required to have like signs and $y_0 > 0$. Our trajectory will accordingly assume the simpler form

$$\begin{aligned} x(s) &= \mu \cos\left(e^s \arccos \frac{x_0}{\mu}\right), \\ y(s) &= \mu\left(\arccos \frac{x_0}{\mu}\right) e^s \sin\left(e^s \arccos \frac{x_0}{\mu}\right), \\ z(s) &= \mu\left(\arccos \frac{x_0}{\mu}\right)^2 e^{2s} \cos\left(e^s \arccos \frac{x_0}{\mu}\right). \end{aligned} \tag{7}$$

The presence of chaotic behaviour in these trajectories can be readily exhibited by introducing a time delay of $\log 2$, where we see that

$$\begin{aligned} x(s + \log 2) &= \mu \cos\left(2e^s \arccos \frac{x_0}{\mu}\right) \\ &= \mu\left(2\cos^2\left(e^s \arccos \frac{x_0}{\mu}\right) - 1\right) = \frac{2}{\mu}(x(s))^2 - \mu, \\ y(s + \log 2) &= 2\mu\left(\arccos \frac{x_0}{\mu}\right) e^s \sin\left(2e^s \arccos \frac{x_0}{\mu}\right) \\ &= 4\mu\left(\arccos \frac{x_0}{\mu}\right) e^s \sin\left(e^s \arccos \frac{x_0}{\mu}\right) \cos\left(e^s \arccos \frac{x_0}{\mu}\right) \\ &= \frac{4}{\mu} x(s) y(s), \\ z(s + \log 2) &= 4\mu\left(\arccos \frac{x_0}{\mu}\right)^2 e^{2s} \cos\left(2e^s \arccos \frac{x_0}{\mu}\right) \\ &= 4\mu\left(\arccos \frac{x_0}{\mu}\right)^2 e^{2s}\left(\cos^2\left(e^s \arccos \frac{x_0}{\mu}\right) - \sin^2\left(e^s \arccos \frac{x_0}{\mu}\right)\right) \\ &= \frac{4}{\mu}\left(x(s) z(s) - (y(s))^2\right). \end{aligned}$$

This means, in particular, that if we put $f(x) = (2/\mu)x^2 - \mu$ and $x(s) = x$, then we will have $x(s + \log 2) = f(x)$, and further iteration of this relation clearly leads to

$$x(s + n \log 2) = f^n(x)$$

for every positive integer n and $s \geq 0$. Therefore, if time s is measured in units of $\log 2$, then the x-coordinates of our trajectory will exhibit the sort of chaotic behaviour that was discussed in earlier sections.

Furthermore, we see that the mapping G from 3-space into itself defined by

$$G(x, y, z) = \left(\frac{2}{\mu}x^2 - \mu, \frac{4}{\mu}xy, \frac{4}{\mu}(xz - y^2)\right)$$

can be embedded in the continuously evolving system of trajectories (7) so long as the initial point (x_0, y_0, z_0) satisfies the conditions of (6). As μ runs over the

positive real numbers, the nature of the mapping G will remain essentially unchanged, for we have

$$\frac{2}{\mu}G\left(\frac{\mu}{2}x, \frac{\mu}{2}y, \frac{\mu}{2}z\right) = \frac{2}{\mu}\left(\frac{2}{\mu}\frac{\mu^2}{4}x^2 - \mu, \frac{4}{\mu}\frac{\mu^2}{4}xy, \frac{4}{\mu}\frac{\mu^2}{4}(xz - y^2)\right)$$
$$= (x^2 - 2, 2xy, 2(xz - y^2)).$$

Evidently the introduction into (7) of a time delay equal to $\log k$ for some positive integer k would result in a mapping G from 3-space into itself that was defined in terms of the Chebyshev polynomials.

4. Orbits. The behaviour of the iterates $f^n(2\cos u)$ for $0 \le u \le \pi$ of $f(z) = z^2 - 2$ as n runs over the positive integers can be analyzed most conveniently by putting $r = u/\pi$ so that $0 \le r \le 1$. The fixed points of f evidently correspond to $r = 0$, $u = 0$ and $r = 2/3$, $u = 2\pi/3$. From the general formula

$$f^n(2\cos \pi r) = 2\cos(2^n \pi r)$$

we conclude that the behaviour of $2^n r$ can be seen to best advantage when we represent r by means of its dyadic expansion

$$r = \frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_k}{2^k} + \cdots = .a_1 a_2 \ldots a_k \ldots,$$

where each a_k is either 0 or 1. Our aim is to determine necessary and sufficient conditions in terms of r for the iterates of $f^n(2\cos \pi r)$ to be absorbed into one or other of the fixed points. From the elementary formula

$$\cos 2\pi r = \cos 2\pi(1 - r)$$

we infer that if $r < 1/2$, then we can compute the next iterate $\cos 2\pi r$ according to the formula

$$\cos 2\pi r = \cos(\pi T(r)),$$

where $T(r)$ is the shift operator defined to be the fractional part of $2r$, that is to say,

$$T(r) = T(.a_1 a_2 \ldots) = .a_2 a_3 \ldots .$$

But if $r > 1/2$, then we must define $T(r)$ to be $1 -$ fractional part of $2r$ according to the formula

$$T(r) = T(.a_1 a_2 \ldots) = .a_2' a_3' \ldots,$$

where $a_k' = 1 - a_k$ for each index k and compute

$$\cos 2\pi r = \cos(\pi T(r)).$$

These definitions of $T(r)$ can be compacted into the single formula

$$T(.a_1 a_2 \ldots) = \begin{cases} .a_2 a_3 \ldots & \text{in case } a_1 = 0, \\ .a_2' a_3' \ldots & \text{in case } a_1 = 1. \end{cases}$$

These two definitions clearly agree in the ambiguous cases

$$T(.0111\ldots) = .111\ldots, \qquad T(.111\ldots) = .000\ldots .$$

The fixed points

$$r = \tfrac{2}{3} = .101010\ldots, \qquad r = 0 = .000\ldots$$

evidently satisfy the relation $T(r) = r$. The relation

$$T(.010101\ldots) = .101010\ldots$$

reveals that the iterates of

$$r = \tfrac{1}{3} = .010101\ldots$$

are quickly absorbed into the fixed point $r = 2/3$. From these observations we can easily deduce the following results.

A necessary and sufficient condition for the iterates $T^n(r)$ of r to be absorbed eventually into the fixed point $r = 2/3$ is that the tail of the dyadic expansion $r = .a_1a_2\ldots$ should be $\ldots 010101\ldots$. Likewise, a necessary and sufficient condition for the iterates $T^n(r)$ of r to be absorbed eventually into the fixed point $r = 0$ is that the tail of r should be $\ldots 000\ldots$, that is to say, r should have a finite dyadic expansion.

As another example let us take $r = .100100\ldots$ and calculate $T(r) = .110110\ldots$, $T(T(r)) = T^2(r) = .010010\ldots, T^3(r) = .100100\ldots$. The iterates of such an r are evidently periodic, and our aim will be to show that *a necessary and sufficient condition for the iterates $T^n(r)$ of r to be eventually periodic is that r should have a dyadic expansion with periodic tail having period p.*

We have already dealt with the cases $p \leq 2$ where the iterates are eventually absorbed into either $r = 0$ or $r = 2/3$. The set of iterates $T^n(r)$ as n runs over the positive integers we shall call the orbit of r and denote it by $O(r)$. The example $r = .011011\ldots$ shows us that the orbit $O(r)$ need not include r in its periodic oscillations, for $T(r) = .110110\ldots$, and the orbit of $T(r)$ has already been charted. Furthermore, the period of $T^n(r)$ as a function of n need not coincide with the period of the dyadic expansion, as shown by the example $r = .11001100\ldots, T(r) = .01100110\ldots, T^2(r) = .11001100\ldots = r$. Nevertheless, we shall verify that the period of $T^n(r)$ as a function of n can not exceed the period of the dyadic expansion. Indeed, the action of the mapping T consists of a truncation of the initial digit and, in some cases, a reversal of all remaining digits. Now the reversals will occur at the first appearance of a 1 in the dyadic expansion, and thereafter at each digit a_k such that $a_{k-1} \neq a_k$. If p is the period of the dyadic expansion, then we have $a_{k+p} = a_k$ for all sufficiently large indices k. If a reversal occurs at a_k, then another must occur at a_{k+p}, provided that a_k is not the first 1 in the expansion. If a_k is the first 1 in the dyadic expansion, then another reversal need not occur at a_{k+p}, as shown by the orbit

$$O(.101101\ldots) = \{.100100\ldots, .110110\ldots, .010010\ldots\}.$$

Evidently an even number of shifts and reversals is equivalent to an even number of shifts with no reversals, and if there were an even number of reversals within one complete period of the dyadic expansion, then the values of $T^n(r)$ would clearly repeat with the same period p. Moreover, if a reversal occurs at a_k and at a_{k+p}, then the digits between them must change at least once, for otherwise the period of the expansion would be 1. Since $a_k = a_{k+p}$, there must be an even number of changes of digit between a_k and a_{k+p}, hence, an even number of reversals, so the resulting action from a_k to a_{k+p} is simply an extended shift, and $T^n(r)$ will eventually be periodic with period p. A well-known elementary argument shows us that the smallest period of $T^n(r)$ as a function of n has to be a divisor of p. The converse of this result, that if $T^n(r)$ is eventually periodic, then

the dyadic expansion of r must be periodic as well, is easily verified. We shall announce this result in the following theorem.

THEOREM 3. *If the dyadic expansion of r is eventually periodic with period p, then $T^n(r)$ will eventually be a periodic function of n, and its period will be a divisor of p. Conversely, if $T^n(r)$ is eventually periodic, then the dyadic expansion of r will also be eventually periodic.*

We shall now turn our attention to orbits $O(r)$ which are infinite. Consider, for example, the orbit $O(r)$ where $r = .1010010001\ldots$. Evidently $O(r)$ is infinite, and its limit points are $1, 1/2, \ldots, 1/2^n, \ldots, 0$, so $O(r)$ is nowhere dense in the closed interval $[0, 1]$.

THEOREM 4. *Given any point r_0 in the interval* $[0, 1]$, *we can find a point r arbitrarily close to r_0 such that $O(r)$ is dense in* $[0, 1]$.

Proof. Let s_0 have a finite dyadic expansion which agrees with sufficiently many of the initial digits of r_0 so that s_0 will lie as close to r_0 as desired. We shall then define a sequence S of finite dyadic expansions as follows:

$$s_1 = .0, \quad s_2 = .1, \quad s_3 = .00, \quad s_4 = .01, \quad s_5 = .10, \quad s_6 = .11, \ldots,$$

according to the obvious pattern wherein all dyadic expansions of length $n - 1$ will have appeared by the time we reach s_{2^n-2}. We now proceed to define the dyadic expansion of r inductively. The initial segment of r will be s_0, and the subsequent segments of r will be either the members s_n of the sequence S in their proper order, or else their reversals s_n', according to the following criterion. Suppose that we have already defined the first n segments which form the first L digits of r. We shall then define the next segment of r to be s_n in case there was an even number of reversals in the first L digits; otherwise, we define the next segment of r to be s_n', where the prime denotes as usual the result of reversing all the dyadic digits of s_n. The infinite dyadic expansion which has thus been formed inductively will be denoted by r, and it is apparent that r lies as close to r_0 as desired. Furthermore, if r_1 is any point in the interval $[0, 1]$ and ε is any small positive number, then we can find a member s_N of S with length M such that $1/2^M < \varepsilon/2$ and $|s_N - r_1| < \varepsilon/2$. From this it follows that some member s of $O(r)$ will have s_N as its initial segment, so

$$|s - r_1| \leq |s - s_N| + |s_N - r_1| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, r_1 must be a limit point of the orbit $O(r)$, and $O(r)$ must be dense in $[0, 1]$.

It is quite clear from the method of proof of Theorem 4 that we can insert an arbitrary initial segment into the dyadic expansion of those points r which we constructed there so as to produce an expansion of r with the property that the resulting orbit $O(r)$ will eventually be periodic. From this it follows that the set of points r in $[0, 1]$ such that $T^n(r)$ is eventually periodic must be dense in $[0, 1]$. Indeed, the set of points r such that $T^n(r)$ is eventually periodic with period $p \geq 3$ must be dense in $[0, 1]$. Finally, it was previously observed that the set of points r such that $T^n(r)$ is eventually absorbed into $r = 0$ is a dense subset of $[0, 1]$, a phenomenon that was noticed by Julia [**11**, p. 186]. The same observation applies to the set of points r such that $T^n(r)$ is eventually absorbed into $r = 2/3$. We shall summarize these observations in the following theorem.

THEOREM 5. *The following sets of points r are dense in the interval* [0, 1]:
(i) *all points r such that* $T^n(r)$ *is eventually absorbed into* $r = 0$ *or* $r = 2/3$;
(ii) *all points r such that* $T^n(r)$ *is eventually periodic with period p*;
(iii) *all points r such that* $O(r)$ *is dense in* [0, 1].

COROLLARY. *If* $f(z) = z^2 - 2$, *then the following sets of points z on the segment* $[-2, +2]$ *of the real axis are dense in that segment*:
(i) *all points z such that* $f^n(z)$ *is eventually absorbed into* $z = 2$ *or* $z = -1$;
(ii) *all points z such that* $f^n(z)$ *is eventually a periodic function of n with period p*;
(iii) *all points z such that* $f^n(z)$ *is dense in* $[-2, +2]$.

Results for the functions f_k of Theorem 2 that are entirely similar to those in the above Corollary can easily be formulated.

5. Further duplication formulas. The initial source of our analytical formulas for the description of chaotic behaviour was the duplication formula for the cosine, which was then extended to higher replications by means of the Chebyshev polynomials. In the present section we shall sample two further duplication formulas and draw conclusions similar to those in earlier sections. A historical connection between the duplication formula for cot z and Cayley's use of Newton's approximation method to solve $z^2 + 1 = 0$ has been pointed out by Mandelbrot. We shall consider first the elementary formula

$$\tanh 2u = \frac{2\tanh u}{1 + \tanh^2 u}.$$

If we put

$$f(z) = \frac{2z}{1 + z^2}, \qquad h(u) = \tanh u, \qquad g(u) = 2u,$$

then clearly $fh(u) = hg(u)$. Iteration of this relation will then yield $f^n h = hg^n$ for each positive integer n, and we have thus arrived at the formula

$$f^n(\tanh u) = \tanh 2^n u$$

for each complex number $u = v + wi$. For real values of u we have $-1 < z = \tanh u < +1$, so the limit as n tends to ∞ of $f^n(z) = \tanh 2^n u$ will be $+1$ in case $u > 0$, 0 in case $u = 0$, and -1 in case $u < 0$. From the relation

$$|\tanh 2^n u|^2 = |\tanh(2^n v + 2^n wi)|^2 = \frac{\sinh^2 2^n v + \sin^2 2^n w}{\cosh^2 2^n v - \cos^2 2^n w},$$

we infer that $f^n(z)$ will tend to a point on the unit circle as n tends to ∞ in case $v \neq 0$. The relation

$$\arg\tanh 2^n u = \arctan\frac{\sin 2^{n+1} w}{\sinh 2^{n+1} v}$$

implies that $\arg f^n(z)$ will tend to either 0 or π as n tends to ∞, and from considerations of continuity we infer that $f^n(z)$ will tend to $+1$ in case $v > 0$ and to -1 in case $v < 0$. A cursory examination of the mapping

$$z = \tanh u = \frac{e^{2u} - 1}{e^{2u} + 1}$$

will serve to remind us that the region $\operatorname{Re} u = v > 0$ is mapped to $\operatorname{Im} z > 0$, while $\operatorname{Re} u = v < 0$ is mapped to $\operatorname{Im} z < 0$. In case $v = 0$ we have

$$|\tanh 2^n u|^2 = \frac{\sin^2 2^n w}{1 - \sin^2 2^n w} = \tan^2 2^n w,$$

$$\arg \tanh 2^n u = \pm \frac{\pi}{2}$$

provided that $w \neq 0$, and we infer that $f^n(z)$ must lie on the imaginary axis. Now the behaviour of $\tan^2 2^n w$ can be analyzed along the same lines that were followed in deriving Theorem 5 as soon as we put $w = (\pi/2)r$, for then we can write

$$\tan^2 2\frac{\pi}{2}(1 - r) = \tan^2 2\frac{\pi}{2}r$$

for all real values of r. Accordingly, we can carry over the results of Theorem 5 *mutatis mutandis* to the present situation and conclude that the following sets of points will be dense on the imaginary axis in the u-plane: the set of points wi such that $\tan^2 2^n w$ is eventually periodic; (ii) the set of points wi such that $\tan^2 2^n w$ is eventually infinite, that is to say, $\tan^2 2^n w = \infty$ for some positive integer n, namely, those points $wi = (2m - 1)2^{-n-1}\pi i$ for any choice of integers m, n with $n \geq 1$; (iii) the set of points wi such that $\tan^2 2^n w$ is dense on the real axis. We shall summarize these results in the following theorem.

THEOREM 6. *If*

$$f(z) = \frac{2z}{1 + z^2}$$

for all points in the complex z-plane, then $f^n(\tanh u) = \tanh 2^n u$ *for all points in the complex u-plane and all positive integers n. As n tends to* ∞, $f^n(z)$ *will converge to* $+1$ *in case* $\operatorname{Re} z > 0$, *to* -1 *in case* $\operatorname{Re} z < 0$, *and to* 0 *in case* $z = 0$. *If* $\operatorname{Re} z = 0$, *then the following sets will be dense in the imaginary axis of the z-plane*: (i) *the set of points z such that* $f^n(z)$ *is eventually a periodic function of n*; (ii) *the set of points z such that* $f^n(z)$ *is eventually infinite*; (iii) *the set of points z such that* $f^n(z)$ *is dense in the imaginary axis.*

COROLLARY. *With the same hypotheses as in Theorem* 6, *the set of points z such that* $f^n(z)$ *remains bounded as n tends to* ∞ *is the complement in the complex z-plane of a dense subset of the imaginary axis.*

6. Higher rational functions. The simple cases of chaotic behaviour which have been discussed in previous sections have barely scratched the surface of the general problems of a topological nature that are to be encountered here. We have already seen examples for certain functions f of the sets of points z in the complex plane where the iterates $f^n(z)$ as n runs over the positive integers are (i) absorbed into the fixed points of f (Theorem 5); (ii) periodic (Theorems 5 and 6); (iii) dense (Theorems 5 and 6); (iv) infinite (Theorem 6). In all of the examples presented so far, these points z were dense in one-dimensional subsets of the complex plane, namely, the real and imaginary axes. In order to illustrate the range of topological possibilities that these sets of points z may have for other rational functions f, we have chosen for our final example a duplication formula from the theory of Jacobian elliptic functions (see [**19**, pp. 491–535]). The duplica-

tion formula for the elliptic function cn u asserts that

$$\operatorname{cn} 2u = \frac{\operatorname{cn}^2 u - \operatorname{sn}^2 u \operatorname{dn}^2 u}{1 - k^2 \operatorname{sn}^4 u} = \frac{k^2(1 - \operatorname{cn}^2 u)^2 - 2(1 - \operatorname{cn}^2 u) + 1}{1 - k^2(1 - \operatorname{cn}^2 u)^2},$$

where k is a positive real number. If we put $z = \operatorname{cn} u$ and

$$F(z) = \frac{k^2(1 - z^2)^2 - 2(1 - z^2) + 1}{1 - k^2(1 - z^2)^2}$$

for all points in the complex z-plane, then it follows that $F(\operatorname{cn} u) = \operatorname{cn} 2u$ and generally that $F^n(\operatorname{cn} u) = \operatorname{cn} 2^n u$ for each positive integer n. Although the behaviour of the iterates $F^n(z)$ of this particular function has not yet excited much attention, it can evidently be deduced in some detail from the poles, zeros, and doubly periodic nature of cn u.

We shall pursue a slightly different course and remark that $F(z)$ can be simplified to some extent by means of elementary substitutions. If we put

$$Z = \frac{k^2(1 - z^2)^2 - 2(1 - z^2) + 1}{1 - k^2(1 - z^2)^2} = F(z)$$

and $t = k(1 - z^2)$, then we will have

$$Z = \frac{t^2 - \frac{2}{k}t + 1}{1 - t^2}.$$

In order to complete our substitutions, we have to put

$$\begin{aligned} T = k(1 - Z^2) &= k(1 - Z)(1 + Z) \\ &= k\frac{\left(\frac{2}{k}t - 2t^2\right)\left(2 - \frac{2}{k}t\right)}{(1 - t^2)^2} \\ &= \frac{4t(1 - kt)\left(1 - \frac{1}{k}t\right)}{(1 - t^2)^2} \\ &\doteq f(t). \end{aligned}$$

Our duplication formula can now be transferred to the function $f(t)$ defined at all points in the complex t-plane. If we recall that

$$t = k(1 - \operatorname{cn}^2 u) = k \operatorname{sn}^2 u,$$

then our previous relation

$$Z = F(z) = F(\operatorname{cn} u) = \operatorname{cn} 2u$$

now assumes the form

$$\begin{aligned} T = f(t) &= f(k \operatorname{sn}^2 u) \\ &= k(1 - Z^2) = k(1 - \operatorname{cn}^2 2u) \\ &= k \operatorname{sn}^2 2u. \end{aligned}$$

Iteration of our new formula $f(k\,\mathrm{sn}^2 u) = k\,\mathrm{sn}^2 2u$ will evidently lead to the result quoted in the following theorem.

THEOREM 7. *If*

$$f(t) = \frac{4t(1-kt)\left(1-\frac{1}{k}t\right)}{(1-t^2)^2}$$

for all points in the complex t-plane, then

$$f^n(k\,\mathrm{sn}^2 u) = k\,\mathrm{sn}^2(2^n u)$$

for each positive integer n.

THEOREM 8. *With the same hypotheses as in Theorem* 7, *if* $0 < k < 1$, *then the following sets of points are dense in the complex t-plane*: (i) *the set of points t such that* $f^n(t)$ *is eventually a periodic function of n*; (ii) *the set of points t such that* $f^n(t)$ *is eventually infinite*; (iii) *the set of points t such that* $f^n(t)$ *is dense in the complex plane.*

Proof. From [**19**] we know that $\mathrm{sn}\,u$ is doubly periodic with periods $4K$ and $2K'i$, where K, K' can be expressed analytically in terms of k, and that if $0 < k < 1$, then K, K' will be real. Furthermore, $\mathrm{sn}\,u$ will be real valued in case u is real and pure imaginary in case u is pure imaginary, while $\mathrm{sn}\,u$ is an odd function. If we put

$$u = v + wi = 2Kr + K'si,$$

then it follows immediately that

$$\begin{aligned}\mathrm{sn}^2 2(2Kr + K'si) &= \mathrm{sn}^2 2(2K(1-r) + K'(1-s)i)\\ &= \mathrm{sn}^2 2(2K(1-r) - K'si) = \mathrm{sn}^2\overline{2(2K(1-r) + K'si)}\\ &= \mathrm{sn}^2\overline{2(2Kr + K'(1-s)i)}.\end{aligned}$$

From Theorem 7 we have

$$f^n\big(k\,\mathrm{sn}^2(2Kr + K'si)\big) = k\,\mathrm{sn}^2 2^n(2Kr + K'si),$$

and if we restrict r, s so as to lie in the interval $[0, 1]$ and represent them by means of their dyadic expansions, then we can write, as in the proof of Theorem 3,

$$\mathrm{sn}^2 2(2Kr + K'si) = \mathrm{sn}^2 T_0(2KT_1(r) + K'T_2(s)i), \tag{8}$$

where T_1, T_2 are composed of shifts and reversals of the respective dyadic expansions, just as in the proof of Theorem 3, and T_0 is the identity mapping in case r, s both lie in either of the intervals $[0, 1/2]$ or $[1/2, 1]$, but T_0 is the conjugation mapping in all other cases.

From these definitions of T_0, T_1, T_2 it follows that $f(t)$ has two fixed points t corresponding to $(r, s) = (0, 0), (2/3, 2/3)$ and two points corresponding to $(r, s) = (0, 2/3), (2/3, 0)$. We can describe the relation between T_0, T_1, T_2 simply by saying that T_0 will be the identity mapping in case both or neither of T_1, T_2 execute a reversal, but T_0 will be the conjugation mapping in case just one of T_1, T_2 executes a reversal. From the relation

$$\mathrm{sn}^2(v + K'i) = \mathrm{sn}^2(v - K'i) = \mathrm{sn}^2\overline{v + K'i}$$

we infer that the alternative definitions in (8) are consistent in the ambiguous cases $r = 1/2, 1$ and $s = 1/2, 1$. Let

$$r = .a_1a_2a_3\ldots, \qquad s = .b_1b_2b_3\ldots$$

denote the dyadic expansions of r, s and suppose now that these expansions are eventually periodic with periods p, q respectively. We have already noticed in the proof of Theorem 3 that there will be an even number of reversals between a_k and a_{k+p}, as well as between b_k and b_{k+q}. If m denotes the least common multiple of p, q and A is the set of indices in a segment of length m where T_1 executes a reversal, while B is the corresponding set of indices in the same segment where T_2 executes a reversal, then we know from the proof of Theorem 3 that A, B will both contain an even number of indices. Furthermore, T_0 will execute a conjugation at all indices in the symmetric difference $D = (A \cup B) - (A \cap B)$, and D must clearly contain an even number of indices regardless of whether $A \cap B$ contains an odd or an even number of indices. Thus, the function $T(n)$ defined by

$$T(n) = T_0^n(2KT_1^n(r) + K'T_2^n(s)i)$$

must eventually be periodic with period m or less. As in the proof of Theorem 3, we readily infer that the period $\bar{p}$ of $T(n)$ must be a divisor of m. The converse, that if $T(n)$ is eventually periodic, then the dyadic expansions must also be periodic, is easily verified.

Now $\operatorname{sn} u$ has simple poles at the points $K'i$ and $2K + K'i$, so it will also have poles at the points

$$4MK + (2N + 1)K'i, \quad 2(2M + 1)K + (2N + 1)K'i$$

for all pairs of integers M, N. By making suitable choices for M, N, and $n \geq 1$ we can fill the square $0 \leq r \leq 1$, $0 \leq s \leq 1$ with a dense set of points having one of the forms

$$\frac{4MK + (2N + 1)K'i}{2^n}, \qquad \frac{2(2M + 1)K + (2N + 1)K'i}{2^n}.$$

If $u = 2Kr + K'si$ is any one of the corresponding points in the u-plane, then $\operatorname{sn}^2(2^n u)$ will evidently be infinite, and we only have now to verify that the mapping $t = k \operatorname{sn}^2 u$ sends the rectangle with corners $0, 2K, 2K + K'i, K'i$ in the u-plane onto a dense subset of the t-plane. To this end we shall use the addition formula for $\operatorname{sn} u$ and write

$$\operatorname{sn} u = \operatorname{sn}(v + wi) = \frac{\operatorname{sn} v \operatorname{dn}' w + i \operatorname{sn}' w \operatorname{cn}' w \operatorname{cn} v \operatorname{dn} v}{\operatorname{cn}'^2 w + k^2 \operatorname{sn}^2 v \operatorname{sn}'^2 w},$$

where we have written $\operatorname{sn}' w$ for $\operatorname{sn}(w, k')$ and likewise for $\operatorname{cn}' w$ and $\operatorname{dn}' w$, while k' denotes the complementary modulus defined by $k^2 + k'^2 = 1$. From the table of values

v	0	K	$2K$	w	0	K'
$\operatorname{sn} v$	0	1	0	$\operatorname{sn}' w$	0	1
$\operatorname{cn} v$	1	0	-1	$\operatorname{cn}' w$	1	0
$\operatorname{dn} v$	1	k'	1	$\operatorname{dn}' w$	1	k

we can trace out the image in the t-plane of the boundary of our rectangle in the u-plane when the latter is traversed in the counterclockwise direction beginning at

$u = 0$. Along the edge $w = 0$ we have $t = k \operatorname{sn}^2 v$, and the image point will trace out the slit from $t = 0$ to $t = k$ along the positive real axis. Along the edge $v = 2K$ we have

$$t = k\left(-i\frac{\operatorname{sn}' w \operatorname{cn}' w}{\operatorname{cn}'^2 w}\right)^2 = -k\frac{\operatorname{sn}'^2 w}{\operatorname{cn}'^2 w},$$

and the image point will trace out the negative real axis from $t = 0$ to $t = -\infty$. Along the edge $w = K'$ we have

$$t = k\left(\frac{k \operatorname{sn} v}{k^2 \operatorname{sn}^2 v}\right)^2 = \frac{1}{k \operatorname{sn}^2 v},$$

and the image point will trace out the slit on the real axis from $t = 1/k$ to $t = +\infty$. Finally, along the edge $v = 0$ we have

$$t = k\left(\frac{i \operatorname{sn}' w \operatorname{cn}' w}{\operatorname{cn}'^2 w}\right)^2 = -k\frac{\operatorname{sn}'^2 w}{\operatorname{cn}'^2 w},$$

and the image point will again trace out the negative real axis but in the opposite direction from $t = -\infty$ to $t = 0$. Consequently, the image in the t-plane of the interior of our rectangle in the u-plane will be the complement in the t-plane of the segments on the real axis from $t = 1/k$ to $t = +\infty$ and from $t = k$ to $t = -\infty$, and we will thus have achieved our construction of a dense set of points in the t-plane such that $f^n(t)$ is infinite for some positive integer n.

We shall employ the construction used in the proof of Theorem 4, suitably modified so as to fit the present situation, for constructing points (r, s) such that the values of $T(n)$ will form a dense subset of the fundamental rectangle with corners at $u = 0, 2K, 2K + K'i, K'i$ as n runs over the positive integers. If the action of the mapping T_0 is ignored for the moment, then the procedure for generating a dense subset of lattice points in our fundamental rectangle could be accomplished in the following manner. At each stage of the process described in the proof of Theorem 4 where we have to assign the list of all dyadic segments of a given length, we should require r to repeat a given segment on that list while s runs over the entire list, then make r repeat the next segment on the list while s runs over the entire list again, and so on in lexicographical order. However, the action of T_0 will intervene in this process whenever one of r, s executes a reversal while the other does not, and the result of this intervention is the conjugate of the value of $T(n)$ that we anticipated. It is an easy matter to identity those regions in the fundamental rectangle where no conjugations will occur during the first n iterations of T_0. Indeed, we have already observed that T_0 is the identity mapping in case (r, s) lies in

$$R_1 = \left[0, \tfrac{1}{2}\right]^2 \cup \left[\tfrac{1}{2}, 1\right]^2,$$

where we have adopted the convention of writing I^2 for the cartesian product $I \times I$ of any subset I of the real axis with itself. Evidently T_0^2 will be the identity mapping in case (r, s) lies in

$$R_2 = \left(\left[0, \tfrac{1}{4}\right] \cup \left[\tfrac{1}{2}, \tfrac{3}{4}\right]\right)^2 \cup \left(\left[\tfrac{1}{4}, \tfrac{1}{2}\right] \cup \left[\tfrac{3}{4}, 1\right]\right)^2,$$

for if g denotes the mapping defined by $g(u) = 2u$, then $g(R_2)$ clearly lies in the region R obtained by translating R_1 through $2MK$ and $NK'i$ as M and N run over the integers. Generally T_0^n will be the identity mapping in case (r, s) lies in the checkerboard region R_n with squares of side 2^{-n} which comprises one half the

area of $[0, 1]^2$, for $g^n(R_n)$ clearly lies in R. If we observe that all the lattice points represented by the dyadic segments used above to generate a dense subset of the fundamental rectangle will themselves lie in the intersection of all the checkerboards R_n as n runs over the positive integers, it then follows that T_0^n will be the identity mapping for all values of n at each of these lattice points, and our construction of a dense subset is complete. The result that $t = \mathrm{sn}^2 u$ maps the fundamental rectangle onto a dense subset of the complex t-plane, which was derived earlier in this proof, is now required to verify that the set of points t such that $f^n(t)$ forms a dense subset of the complex plane as n runs over the positive integers is itself dense in the complex t-plane. At the same time, this property of $t = \mathrm{sn}^2 u$ implies that the set of points t such that $f^n(t)$ becomes eventually a periodic function of n is itself dense in the complex t-plane.

COROLLARY. *With the same hypotheses as in Theorem* 8, $f^n(t)$ *will eventually be a periodic function of* n *with period* $\bar{p}$ *if, and only if, the dyadic expansions of* r, s *are both periodic with periods* p, q, *respectively, and* $\bar{p}$ *will divide the least common multiple of* p, q.

REFERENCES

1. M. F. Barnsley, J. S. Geronimo and A. N. Harrington, Geometry and combinatorics of Julia sets of real quadratic maps., *J. Statist. Phys.*, 37 (1984) 51–92.
2. M. F. Barnsley and A. N. Harrington, A Mandelbrot set for pairs of linear maps., *Phys. D.*, 15 (1985) 421–432.
3. M. F. Barnsley J. S. Geronimo, and A. N. Harrington, Geometrical and Electrical Properties of Some Julia Sets, Lecture Notes in Pure and Appl. Math., 92. Dekker, New York, 1984.
4. _____, Geometry, electrostatic measure and orthogonal polynomials on Julia sets for polynomials, *Ergodic Theory Dynamical Systems*, 3 (1983) 509–520.
5. Paul Blanchard, Complex analytic dynamics on the Riemann sphere, *Bull. Amer. Math. Soc.*, (N. S.) 11 (1984) 85–141.
6. G. J. Butler, and G. Pianigiani, Periodic points and chaotic functions in the unit interval, *Bull. Austral. Math. Soc.*, 18 (1978) 255–265.
7. P. Collet, and J.-P. Eckmann, On the abundance of aperiodic behavior for maps on the unit interval, *Bull. Amer. Math. Soc.*, (N. S.) 3 (1980) 699–700.
8. _____, Iterative Maps on the Interval as Dynamical Systems, Birkhauser, Boston, 1980.
9. Robert L. Devaney, An Introduction to Chaotic Dynamical Systems, Benjamin Cummings, Menlo Park, Calif., 1986.
10. M. P. Fatou, Sur les équations fonctionnelles, *Bull. Soc. Math. France*, 47 (1919) 161–271. See also ibid. 48 (1920), 33–94, 208–314.
11. G. Julia, Mémoire sur l'itération des fonctions rationnelles, *J. Math. Pures Appl.*, 4 (1918) 47–245.
12. S. Lattès, Sur l'itération des substitutions rationelles et les fonctions de Poincaré, *C. R. Acad. Sci. Paris*, 166 (1918) 26–28.
13. Tien-Yien Li and James A. Yorke, Period three implies chaos, *Amer. Math. Monthly*, 82 (1975) 985–992.
14. Benoit B. Mandelbrot, The Fractal Geometry of Nature, Freeman, New York, 1983.
15. _____, On the quadratic mapping $z \to z^2 - \mu$ for complex μ and z, *Physica*, 7D (1983) 224–239.
16. _____, On the dynamics of iterated maps VIII, Springer Series in Synergetics, 24, Springer-Verlag, New York, 1984.
17. C. Preston, Iterates of Maps on an Interval, Lecture Notes in Mathematics No. 999, Springer-Verlag, New York, 1983.
18. David Whitley, Discrete dynamical systems in dimensions one and two, *Bull. London Math. Soc.*, 15 (1983) 177–217.
19. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, fourth edition, Cambridge University Press, 1973.

Multiple Pythagorean Number Triples

ALBERT FÄSSLER*



ALBERT FÄSSLER obtained his bachelor's degree in Mechanical Engineering at the HTL in Winterthur before he studied Mathematics at the Eidgenössische Technische Hochschule ETH in Zürich (Federal Institute of Technology). After graduation, he worked under Eduard Stiefel at the Seminar für Angewandte Mathematik (Seminar for Applied Mathematics) at the ETH, where he got his Ph.D. and became a lecturer. Together with Stiefel, he wrote a book in the field of applied group theory with the title *Gruppentheoretische Methoden und ihre Anwendung* (Teubner, Stuttgart, 1979). He is now at the HTL Biel/Bienne (an Engineering college).

SUMMARY OF THIS ARTICLE. We focus on the following questions: How many Pythagorean number triples are there, whose triangles have a common hypotenuse, leg-sum, leg-difference, area, perimeter, inradius or leg, and how can they be calculated? How are such triples distributed for small parameters as well as asymptotically? Statistical results generated by computer will be given. Some of the asymptotic statements may appear surprising to readers who are relatively unfamiliar with number theory. A survey of published results and historical remarks are included in this article to make it more enjoyable.

1. Introduction. A triple of strictly positive integers x, y, h satisfying

$$x^2 + y^2 = h^2 \tag{1.1}$$

is called a *Pythagorean number triple*, or more briefly a *Pythagorean triple*.

A Pythagorean triple, denoted by $[x, y, h]$, is called *primitive* if and only if x, y, h are relatively prime, i.e. their gcd is 1; it is sufficient that $\gcd(x, y) = 1$ holds.

If x and y were both odd then (1.1) would imply that $h^2 \equiv 2 \pmod 4$. But this is impossible, because if h is even this implies $h^2 \equiv 0 \pmod 4$ and if h is odd this implies $h^2 \equiv 1 \pmod 4$. Consequently, one leg has to be even and the other has to be odd. The following way of describing primitive Pythagorean triples (ppt) has been known at least since the time of Euclid.

The pairs of integers $[a, b]$ with the properties

$$a > b > 0 \qquad \gcd(a, b) = 1 \qquad a + b \equiv 1 \pmod 2 \tag{1.2}$$

generate the set of all ppt $[x, y, h]$ with x even and y odd by taking

$$x = 2ab \qquad y = a^2 - b^2 \qquad h = a^2 + b^2. \tag{1.3}$$

The above description is a *bijection* between the pairs $[a, b]$ with properties (1.2) and all the ppt $[x, y, h]$ with even x.

If we use any pair of integers a, b with $a > b > 0$ in (1.3), we would get a Pythagorean triple, but it would not be primitive in general. We call a and b the *generators* of the primitive triple $[x, y, h]$.

*Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, and Ingenieurschule Biel, Postfach 1180, CH-2500 Biel/Bienne Switzerland. This work was mainly done when the author was spending his sabbatical in Minneapolis. He wishes to thank the State of Bern, in Switzerland, for its support and the IMA, at the University of Minnesota, for its provision of excellent working conditions and computer facilities.

Examples:

$$[2,1] \leftrightarrow [4,3,5] \qquad [3,2] \leftrightarrow [12,5,13] \qquad [21,4] \leftrightarrow [168,425,457].$$

GEOMETRIZATION. (1.3) *can be considered as the conformal mapping* $z \mapsto z^2$ *with* $z = a + ib$:

$$(a + ib)^2 = (a^2 - b^2) + i \cdot 2ab. \tag{1.4}$$

Indeed, $y = \mathrm{Re}(z^2)$ and $x = \mathrm{Im}(z^2)$. Introducing coordinate axes as shown in FIGURE 1, the mapping $z \mapsto z^2$ has the following properties when discussed for the entire (a, b)-plane:

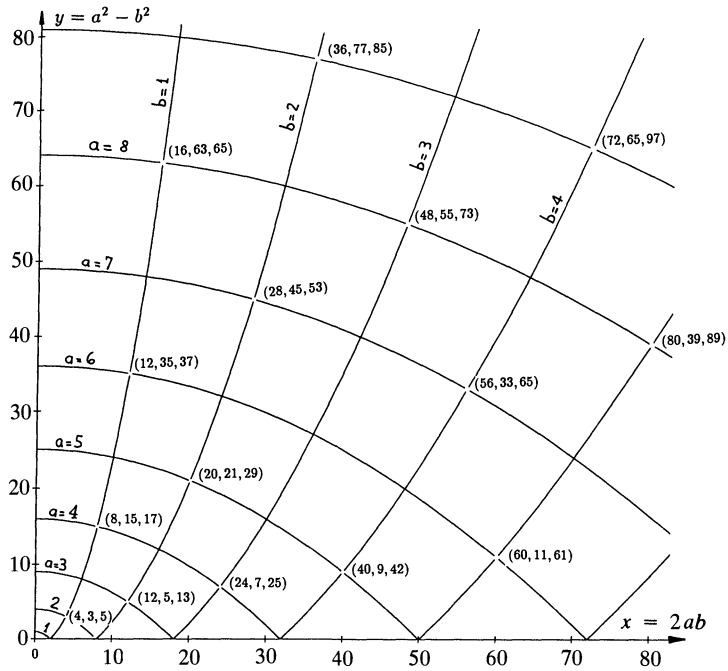


FIG. 1

For $a = a_0 = \text{const}$, we get $y = a_0^2 - b^2$ and $x = 2a_0 b$. For $a_0 \neq 0$ this implies $b = x/2a_0$ so that

$$y = -\frac{x^2}{4a_0^2} + a_0^2. \tag{1.5}$$

As a result, the images of horizontal lines are parabolas which are symmetric about the y-axis, and which open downward, with their common focus at the origin O. The exceptional case of $a_0 = 0$ gives us the negative y-axis as the image.

Similarly, the images of vertical lines $b = b_0 = const$ are described by

$$y = \frac{x^2}{4b_0^2} - b_0^2, \qquad b_0 \neq 0 \tag{1.6}$$

and they are parabolas that open upwards.

Furthermore, $y = 0$ implies $x^2 = 4a_0^4$ or $x^2 = 4b_0^4$, respectively. Consequently, parabolas of the two orthogonal families have a point on the positive x-axis in common if and only if $a_0 = b_0$.

FIGURE 2 shows the images (x, y) of all points (a, b) with $a > b > 0$, $a = 1, 2, 3, \ldots, 42$ and $b = 1, 2, 3, \ldots, 30$. In particular, the computer generated points mark all ppt in the range x and $y \lessapprox 1800$, where the distance from the origin to such a point gives the length of the hypotenuse of the corresponding right triangle. The points marked with an h correspond to primitive multiple triples with a common hypotenuse. Similarly, s stands for a common leg-sum and f for a common area. See the following for details.

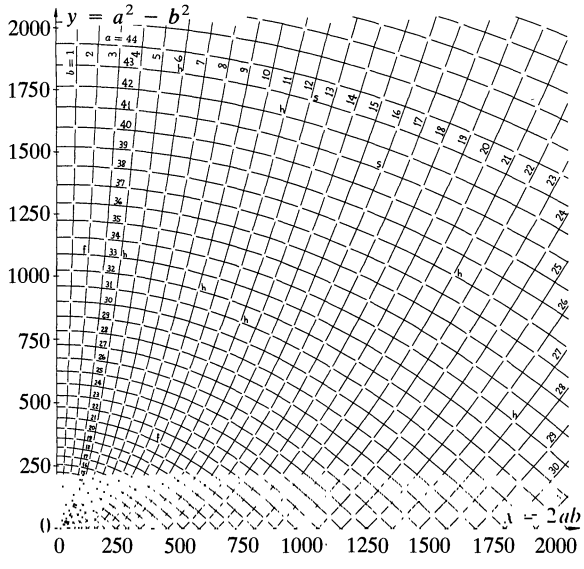


FIG. 2

Note that

$$\text{Number of points on a curve with const } a\text{-value} = \begin{cases} \varphi(a) & \text{if } a \text{ is even} \\ \varphi(a)/2 & \text{if } a \text{ is odd,} \end{cases} \tag{1.7}$$

where $\varphi(a)$ = number of positive integers less than a and relatively prime to a. That is, φ is the Euler-*phi function*. If a is even, (1.7) holds by definition. If a is odd, we use the fact that $gcd(a, k) = gcd(a, a - k)$ holds for every k. But k and $a - k$ have different parity. Therefore, exactly one of a pair $k, a - k$ is even.

If by reflection with respect to the line $x = y$ in FIGURE 2, we map all the points with $y > x$ into the domain $x > y > 0$, we get FIGURE 3. The density of ppt doubles if we drop the condition that x must be even, however we lose the nice geometrical pattern.

Please note: We always identify triangles that are mirror images of each other.

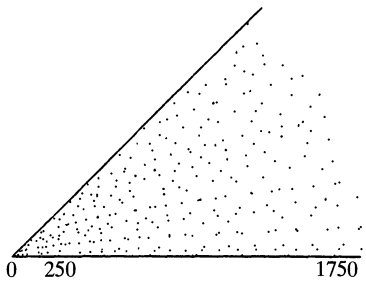


FIG. 3

2. Triples with Common Hypotenuse h. If $\gcd(a, b) = 1$, then any odd prime divisor of $h = a^2 + b^2$ is of the form $4m + 1$. See [**4**, Theorem 13].

Because h is odd, it is necessary that its prime factorization have the form

$$\begin{cases} h = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdots p_n^{\beta_n} \\ \text{with pairwise different primes } p_i \equiv 1 \pmod 4 \text{ and } \beta_i \geqq 1. \end{cases} \tag{2.1}$$

Fermat established that each prime number p of the form $4m + 1$ can be written uniquely as a sum of two squares: $p = a^2 + b^2$, up to a permutation of the squares. We know from the above discussion that a, b have different parity.

When discussing representations of a positive integer as a sum of two squares, we will need the following two identities:

$$\begin{cases} (r^2 + s^2) \cdot (u^2 + v^2) = (ru - sv)^2 + (rv + su)^2 \\ (r^2 + s^2) \cdot (u^2 + v^2) = (ru + sv)^2 + (rv - su)^2. \end{cases} \tag{2.2}$$

Consequently, if we can represent each factor as a sum of two squares, (2.2) gives us their product as a sum of two squares.

Furthermore, let us call the representation of a number as a sum of two squares $a^2 + b^2$ *proper*, if $\gcd(a, b) = 1$. With the knowledge of the representation for the primes p_i in (2.1), the unique proper representations for the prime powers $p_i^{\beta_i}$ as well as the total set of 2^{n-1} different proper representations for h can be generated by the composition (2.2): Each additional prime power $p_i^{\beta_i}$ doubles the number of proper representations. Fermat (1632–1690), Euler (1707–1783), Lagrange (1736–1813) and Legendre (1752–1833) played important roles in the development of these ideas (see [**18**, Chap. II, §VIII and §IX] and [**3**, Chap. VI]). The parity condition is respected by the two compositions (2.2).

In summary, we have the following:

THEOREM 1. *If $h > 1$ has the form* (2.1), *then there are exactly 2^{n-1} different primitive Pythagorean triples with common hypotenuse h, which can be generated by applying* (2.2). *If h is not of the form* (2.1), *then there is no primitive Pythagorean triple.*

Examples: The smallest h with exactly

(a) 1 ppt is $5 = 2^2 + 1^2$ and the triple is $[4, 3, 5]$

(b) 2 ppt is $5 \cdot 13 = 65 = (2^2 + 1^2)(3^2 + 2^2) = (6 \pm 2)^2 + (4 \mp 3)^2 = 8^2 + 1^2 = 4^2 + 7^2$ and the triples are $[16, 63, 65]$ and $[56, 33, 65]$.

(c) 4 ppt is $65 \cdot 17 = 1105$. From $[8, 1], [7, 4]$ for 65, and using $[4, 1]$ for 17, we get:

$$[33, 4] \leftrightarrow [264, 1073, 1105] \qquad [31, 12] \leftrightarrow [744, 817, 1105]$$
$$[32, 9] \leftrightarrow [576, 943, 1105] \qquad [24, 23] \leftrightarrow [1104, 47, 1105]$$

(d) 128 ppt is $5 \cdot 13 \cdot 17 \cdot 29 \cdot 37 \cdot 41 \cdot 53 \cdot 61 = 138,863,452,745$.

The initial elements of the ordered sequence of h-values with

(1) 4 ppt are: 1105 1885 2405 2465 2665 3145 3445 3485 3965 4505 4745 5185
Only the first two, marked with h, are shown in FIGURE 2.

(2) 8 ppt are: 32,045 40,885 45,305 58,565 67,405 69,745 77,285 80,665

DISTRIBUTION OF TRIPLES WITH COMMON HYPOTENUSE.

The asymptotic behavior of the number $\pi(x, m, r)$ of all primes $p \leqq x$ with $p \equiv r \pmod m$ and $\gcd(r, m) = 1$, as $x \to \infty$, the so called primes in *arithmetic progression*, (see [**10**, §4, Ch. III, Th. 3.17]) is given by:

$$\pi(x, m, r) \sim \frac{1}{\varphi(m)} \cdot \frac{x}{\log x}. \tag{2.3}$$

In our case, with the notation $\pi(x)$ for the number of primes $\leqq x$, we get

$$\pi(x, 4, 1) \sim \frac{1}{2} \cdot \frac{x}{\log x} \sim \frac{1}{2} \cdot \pi(x). \tag{2.4}$$

To generate h-values with several ppt, one has to deal with primes of the form $4m + 1$. TABLE 1 shows some values of $\pi(x)$ and $\pi(x, 4, 1)$.

TABLE 1. Distribution of primes.

x	$\pi(x)$	$\pi(x, 4, 1)$
100	25	11
1,000	168	80
2,000	303	147
4,000	550	269
8,000	1007	499
16,000	1862	920
32,000	3432	1705

We let $H_n(x)$ be the number of $h \leqq x$ such that there exist exactly 2^{n-1} different ppt $[x, y, h]$ for each h. Computer calculations resulted in TABLE 2 and FIGURE 4.

TABLE 2. Distribution of multiple ppt with common hypothenuse.

x	$H_2(x)$: 2 triples	$H_3(x)$: 4 triples	$H_3(x)$: 8 triples
10,000	100	10	0
20,000	183	26	0
40,000	338	35	1
100,000	664	102	4
200,000	1,183	170	5
500,000	2,429	363	20
1,000,000	4,184	654	58
2,000,000	7,608	1,251	78
4,000,000	13,236	1,943	174
10,000,000	28,624	4,090	522
20,000,000	51,574	7,284	707
50,000,000	111,773	15,203	1,718
100,000,000	204,891	28,300	3,480
200,000,000	369,064	44,444	6,401
400,000,000	680,850	79,373	11,138
1,000,000,000	1,519,597	161,274	22,640
2,000,000,000	2,803,231	301,286	35,707

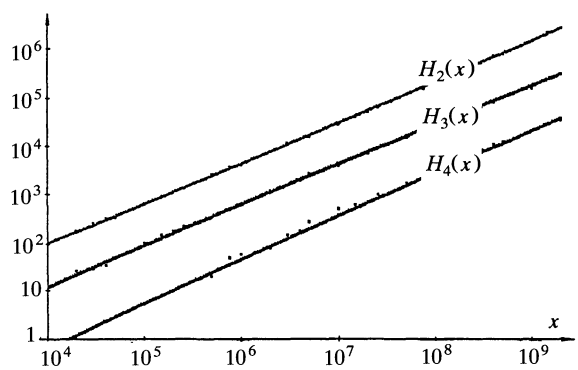


FIG. 4

To discuss the *asymptotic behaviour of* $H_n(x)$, we start with a classical result:

If $\pi_n(x)$ is the number of all positive integers $\leqq x$ of the form (2.1), but *without* the condition $p_i \equiv 1 \pmod 4$, then for each $n = 1, 2, 3, \ldots$ we have

$$\pi_n(x) \sim \frac{1}{(n-1)!} \cdot \frac{x(\log\log x)^{n-1}}{\log x} \qquad (x \to \infty). \tag{2.5}$$

See [**9**, Ch. III, §4, Th. 3.12] or [**4**, Sec. 22.18, Th. 437] (the case $n = 1$ is the prime number theorem).

Because (2.4) holds, we have to introduce a factor 2 for each prime factor of the form $4m + 1$. From (2.5) it follows that

$$H_n(x) \sim \frac{1}{2^n} \cdot \frac{1}{(n-1)!} \cdot \frac{x(\log\log x)^{n-1}}{\log x} \qquad (x \to \infty). \tag{2.6}$$

Furthermore, for each $n = 2, 3, 4, \ldots$,

$$\frac{H_{n-1}(x)}{H_n(x)} \sim \frac{2(n-1)}{\log\log x} \qquad (x \to \infty), \tag{2.7}$$

and, therefore,

$$\lim_{x\to\infty} \frac{H_{n-1}(x)}{H_n(x)} = 0. \tag{2.8}$$

The asymptotic behaviour is completely different from what we have seen in TABLE 2 and FIGURE 4. However, x has to be "huge." The monotonic function on the right-hand side of (2.7) decreases extremely slowly. Its value is 1 at $x = e^{e^{2(n-1)}}$.

REMARK. In [**8**] and [**15**], the asymptotic behaviour of the number of *all ppt* with $h \leqq x$ is given by

$$\sum_{n=1}^{\infty} 2^{n-1} \cdot H_n(x) \sim \frac{x}{2 \cdot \pi}.$$

FORMULA OF LEGENDRE. *For given $h' = \gamma h > 0$, where h is of the form* (2.1) *and γ does not contain divisors of the form $4m + 1$, the total number M of Pythagorean triples that are corresponding to the hypotenuse h' is given by*

$$2M + 1 = (2\beta_1 + 1) \cdot (2\beta_2 + 1) \cdots (2\beta_n + 1). \tag{2.9}$$

This is a straightforward result from Theorem 1. We count all possible ppt, whose hypotenuses are divisors d of h'. The ppt can then be multiplied by h'/d to get the desired triples:

For $n = 1$, we have β_1 triples: all possibilities are $d = p_1^k$ with $1 \leqq k \leqq \beta_1$.

For $n = 2$, we have $2\beta_1\beta_2 + \beta_1 + \beta_2$ triples, because $d = p_1^j p_2^k$ with $j, k \geqq 1$ has two common ppt.

By induction, we arrive at the general case.

SOME CONSEQUENCES.

(1) A Pythagorean triple exists if and only if the hypotenuse has at least one prime factor of the form $4m + 1$. Example: The representation for $10 = 3^2 + 1$ is unique.
(2) If $\gamma > 1$, every triple corresponding to the hypotenuse h' is *not* primitive.
(3) Every triple with the corresponding hypotenuse h' is primitive if and only if $h' = p_1 \cdot p_2 \cdots p_n$ with pairwise different primes $p_i \equiv 1 \pmod 4$.
(4) The special case $n = 1$ in (2.9) implies that $M = \beta_1$: For an arbitrarily chosen integer $M > 0$ there is an infinity of hypotenuses h', each of them corresponding to M different Pythagorean triples. Example: The minimal value for h with three triples is 5^3. The triples are

$$5^2 \cdot [4,3,5] \quad 5 \cdot [24,7,5^2] \quad [336,527,5^3].$$

3. Triples with Common Leg-Sum s. From (1.2), for a given leg-sum we get

$$s = x + y = (a^2 - b^2) + 2ab = (a + b)^2 - 2b^2.$$

With $u = a + b$ and $v = b$, this leads to Pell's equation

$$u^2 - 2v^2 = s \qquad \text{with } \gcd(u, v) = 1. \tag{3.1}$$

Because $a > b$, we have $a + b > 2b$ and

$$u > 2v > 0 \tag{3.2}$$

must also hold.

Therefore, the solutions of (3.1) with the property (3.2) give the solutions of all ppt with leg-sum s, where $a = u - v$, $b = v$.

To deal with Pell's equation, we write

$$u^2 - 2v^2 = (u + v\sqrt{2})(u - v\sqrt{2}) \tag{3.3}$$

and we work in the *integer ring* $\mathbb{Z}[\sqrt{2}]$ consisting of elements $\alpha = u + v\sqrt{2}$ with *norm* $N(\alpha) = u^2 - 2v^2$. With the *conjugate element* $\alpha' = u - v\sqrt{2}$, we obtain

$$N(\alpha) = \alpha \cdot \alpha' = u^2 - 2v^2, \tag{3.4}$$

which is again (3.3). Indeed, solving Pell's equation is the same as finding elements with a given norm s.

Let us multiply the right-hand side of

$$(r^2 - 2s^2)(u^2 - 2v^2) = (r + s\sqrt{2})(r - s\sqrt{2})(u + v\sqrt{2})(u - v\sqrt{2})$$

in two different ways:

(1) first with third, and second with fourth factor;

(2) first with fourth, and second with third factor.

Then we obtain

$$\begin{cases} (r^2 - 2s^2)(u^2 - 2v^2) = (ru + 2sv)^2 - 2 \cdot (rv + su)^2 \\ (r^2 - 2s^2)(u^2 - 2v^2) = (ru - 2sv)^2 - 2 \cdot (rv - su)^2. \end{cases} \tag{3.5}$$

Elements with norm $+1$ or -1, so called *units*, play a significant role. The complete set of units in $\mathbb{Z}[\sqrt{2}]$ is the following: $\pm(1 + \sqrt{2})^m$, $m \in \mathbb{Z}$. Their norms are equal to $(-1)^m$, because $N(\alpha \cdot \beta) = N(\alpha) \cdot N(\beta)$ always holds. If we know one solution (u_0, v_0) of Pell's equation, we can generate recursively an infinite set of solutions by introducing the two units $(1 + \sqrt{2})^2 = 3 + 2\sqrt{2}$ and $(3 + 2\sqrt{2})^{-1} = 3 - 2\sqrt{2}$ with norm $+1$ in (3.5):

$$\begin{cases} u_{k+1} + v_{k+1}\sqrt{2} = (3 + 2\sqrt{2})(u_k + v\sqrt{2}) = (3u_k + 4v_k) + (+2u_k + 3v_k)\sqrt{2} \\ u_{k-1} + v_{k-1}\sqrt{2} = (3 - 2\sqrt{2})(u_k + v\sqrt{2}) = (3u_k - 4v_k) + (-2u_k + 3v_k)\sqrt{2} \end{cases} \tag{3.6}$$

$k = 0, 1, 2, \ldots$.

Note that the indices in the first case are increasing and positive, in the second case decreasing and negative.

Two elements α and β in $\mathbb{Z}[\sqrt{2}]$ are called *associates* if and only if $\alpha = \beta \cdot \varepsilon$ for some unit ε. This is an equivalence relation. All integers $u_j + v_j\sqrt{2}$ generated in (3.6) form a subset of an equivalence class of $\mathbb{Z}[\sqrt{2}]$: See [**5**, Ch. 16, Section 16.5].

The generalized identities

$$\begin{cases} (r^2 + Rs^2)(u^2 + Rv^2) = (ru - Rsv)^2 + R \cdot (rv + su)^2 \\ (r^2 + Rs^2)(u^2 + Rv^2) = (ru + Rsv)^2 + R \cdot (rv - su)^2 \end{cases} \tag{3.7}$$

contain the two special cases (2.2) and (3.5) if we choose $R = 1$ and $R = -2$ respectively. In the previous section, we had the ring $\mathbb{Z}[i]$ of the Gaussian integers with norm $N(u + iv) = |u + iv| = u^2 + v^2$, where $\pm 1, \pm i$ were the only units. The fact that we have an infinity of units makes the leg-sum problem tougher than the hypotenuse problem.

Now we use results from the Gaussian Law of Quadratic Reciprocity and Quadratic Forms. See [**9**, Ch. VI, Th. 111], [**11**, Ch. 2], [**1**, Ch. 4].

LEMMA 1.

(A) *There is a solution of problem* (3.1) *if and only if all prime factors of* s *are of the form* $8m \pm 1$.

(B) *If* $s > 1$ *is a product of* n *different prime powers, each prime being of the form* $8m \pm 1$, *then among the infinity of solutions, there are exactly* 2^{n-1} *pairwise nonassociated solutions of problem* (3.1) *with* $u > 0$ *and* $v > 0$.

Note that 2^{n-1} *is half the number of solutions of the equation* $u^2 - 2 \equiv 0 \pmod{s}$.

The question now is: How many of the infinity of solutions of (3.1) satisfy $u > 2v > 0$?

From

$$\frac{u^2 - 4v^2}{u + 2v} = \frac{2s - u^2}{u + 2v} = u - 2v > 0 \tag{3.8}$$

it follows that there are only a *finite* number of such pairs (u, v), if any. Trivially, if (u, v) is a solution of (3.1), then all four combinations $(\pm u, \pm v)$ are solutions. More is said in

LEMMA 2. *Assumption:* s *has the form mentioned in* (B). *Among the infinity* $(u, v) = (\pm u_k, \pm v_k)$ *of solutions of* (3.1), *where the set* (u_k, v_k), $k \in \mathbb{Z}$ *is described in* (3.6), *there is exactly one which satisfies the condition* $u > 2v > 0$.

Lemma 2 is proved in [**12**]. This article deals with the "Problèmes de Fermat," in which Frenicle was involved too. See also [**18**, Ch. II, §XII].

GENERATING PRIMITIVE PYTHAGOREAN TRIPLES WITH COMMON LEG-SUM. *Using* (3.5) *for the* n *different prime powers of* $s > 1$ *gives* 2^{n-1} *solutions of* (3.1), *if all prime factors of* s *are of the form* $8m \pm 1$. *Iterating, in the sense of* (3.6) *until* $|u| > 2|v|$ *holds, generates the desired ppt with* $a = |u| - |v|$ *and* $b = |v|$.

Examples: The smallest leg-sum with two different ppt is $s = 7 \cdot 17 = 119$. $u^2 - 2v^2 = 7$ has the solution $u = 3$, $v = 1$, and $u^2 - 2v^2 = 17$ has the solution $u = 5$, $v = 2$. Therefore, using (3.5), $u^2 - 2v^2 = 119$ has the solutions $(u, v) = (19, 11)$ and $(11, 1)$. The first does not satisfy the condition $u > 2v$. From the second line of (3.6) we have

$$u = 3 \cdot 19 - 4 \cdot 11 = 13 \qquad v = -2 \cdot 19 + 3 \cdot 11 = -15.$$

One step is sufficient here. With (11, 1) and (13, 5), we obtain the following result: [10, 1] generates [20, 99, 101] and [8, 5] generates [80, 39, 89].

The smallest leg-sum with 4 different ppt is $s = 7 \cdot 17 \cdot 23 = 2737$. Using the two results for u, v above and the fact, that $u^2 - 2v^2 = 23$ has the solution (11, 7) in each case after at most one iteration we get the generators: [37, 36][41, 16][43, 12][47, 6] with the ppt

$$[2664, 73, 2737][1312, 1425, 2737][1032, 1705, 2737][564, 2173, 2737].$$

See also FIGURE 2, where the particular points are marked with s, assuming they appear in the diagram. There are more correspondences to the section dealing with common hypotenuses:

(1) If on the right-hand side of (3.1) the leg-sum $s > 1$ and the prime factors are of the form $8m \pm 1$, then Theorem 1 holds for common s instead of common h, and so does the Legendre formula (2.9).

(2) *Asymptotic behavior*: Using $\pi(x, 8, \pm 1)$ in (2.3), the corresponding definition of $H_n(x)$ (see definition following TABLE 1), denoted by $S_n(x)$ and using the argument following (2.5), we get:

$$\frac{S_{n-1}}{S_n} \sim \frac{4(n-1)}{\log\log x} \qquad n = 2, 3, 4, \ldots \tag{3.9}$$

$$\lim_{x\to\infty} \frac{S_{n-1}}{S_n} = 0 \tag{3.10}$$

similar to (2.7) and (2.8).

4. Triples with Common Leg-difference, Area, Perimeter, Inradius or Leg.

1. *Common Leg-difference d.* With a common leg-difference $d = |x - y|$ and (1.3) we have

$$\begin{aligned} d &= |2ab - (a^2 - b^2)| \\ &= |2a^2 - (a + b)^2| \quad \text{with } a > b > 0, \\ &\quad \gcd(a, b) = 1, \qquad a + b = \text{odd}. \end{aligned} \tag{4.1}$$

Also in this case, we have to deal with problem (3.1), however without the restrictive condition (3.2). So it is not astonishing to get an infinity of solutions, if all prime factors of d are of the form 8 ± 1 and no solution otherwise. In [**20**] we can find the easily verifiable statement:

If $[a, b]$ *generates a ppt with leg-diffeence d then so does* $[a + 2b, b]$.

Therefore, if we know one of the ppt, then an infinity of ppt can be generated.

2. *Common Area f.* Searching for ppt with a common area f leads to the problem of finding f such that the equation

$$\begin{aligned} f &= \frac{1}{2} \cdot xy = ab(a - b)(a + b) \quad \text{with } a > b > 0, \\ &\quad gcd(a, b) = 1, \qquad a + b = \text{odd} \end{aligned} \tag{4.2}$$

has several solutions. This problem is hard to solve. In [**16**], we learn that

The smallest area with a total of
(a) 2 ppt is $2 \cdot 3 \cdot 5 \cdot 7 = 210$, generated by $[6, 1][5, 2]$.
(b) 3 ppt is $210 \cdot 11 \cdot 13 \cdot 19 \cdot 23 = 13123110$, generated by

$$[77, 38][78, 55][138, 5].$$

In [**19**], an infinity of *pairs* of ppt with common area, but no information on more than 2 ppt with common area is given. Despite Euler's efforts on this problem, not much more can be said today. The asymptotic behaviour for the number $T_A(x)$ of *all ppt* with area $\leqq x$, denoted by $T_A(x)$, is given by (see [**19**]):

$$T_A(x) \sim c \cdot \sqrt{x} \quad \text{with } c = \frac{\Gamma(1/4)^2}{\sqrt{2} \cdot \pi^{5/2}} \approx 0.531340. \tag{4.3}$$

We denote by $A_k(x)$ the number of those areas $\leqq x$ which are associated with exactly k ppt. With the aid of the computer, we found for $x = 1{,}588{,}020$: $A_1 = 1976$, $A_2 = 22$, $A_3 = 1$ and $A_4 = A_5 = \cdots = 0$.

In [**6**] and [**14**], an efficient method for generating rational right triangles whose areas are *integers* is given.

3. *Common Perimeter p*

$$p = x + y + h = 2ab + (a^2 - b^2) + (a^2 + b^2) = 2(a + b)a \tag{4.4}$$

with the conditions mentioned in (4.2).

Therefore p = even is the perimeter of a ppt if and only if

$$\frac{p}{2} = u \cdot v \quad \text{with} \quad u > v > 0 \quad \text{and} \quad \gcd(u, v) = 1. \tag{4.5}$$

The generators are $a = u - v$, $b = v$. $u > 2v$ and (4.5) imply the number of isoperimetrical ppt with even perimeter p is equal to the number of odd divisors u of $p/2$ with $\sqrt{p/2} < u < \sqrt{p}$ and $\gcd(u, p/2v) = 1$.

Using Bertrand's postulate, it is proven in [**7**], that for each $m > 0$, there is a perimeter p with at least m isoperimetrical ppt.

In [**2**], a list of 5 perimeters between 14,280 and 103,740 is given ($60 \leqq a \leqq 210$), each one associated with exactly 3 ppt. The asymptotic behaviour for the number T_P of *all ppt* with perimeter $\leqq x$ is given by (see [**8**]):

$$T_P(x) \sim \frac{\log 2}{\pi^2} \cdot x. \tag{4.6}$$

4. *Common Inradius i*

$$i = \frac{2f}{p} = \frac{ab(a + b)(a - b)}{(a + b)a} = b(a - b) \quad \text{with} \quad a - b = \text{odd}. \tag{4.7}$$

As a result of (4.7), i is always an integer, and the problem is simpler than that of the perimeter: If i has the prime factorization

$$i = 2^{\beta} p_1^{\beta_1} \cdot p_2^{\beta_2} \cdots p_n^{\beta_n} \quad \text{with} \quad 2 < p_1 < p_2 < \cdots < p_n,$$
$$\beta \geqq 0, \qquad \beta_i \geqq 1, \tag{4.8}$$

then

the number of ppt with common inradius i is 2^n,

because it is equal to the number of ways to express i as a product of two relatively prime factors. An extensive study of this case can be found in [**17**, Sec. 23.2].

5. *Common Leg.* This problem is left to the reader.

5. Triples with Two Common Items.

THEOREM 3. *There exists no two noncongruent right triangles with two of the following items in common: hypotenuse, leg-sum, leg-difference, area, perimeter, inradius, one leg.*

Note that the theorem is not restricted to integers. Its result is quite obvious: 3 elements of a triangle essentially (with exceptions) determine it.

Proof. Some of the $\binom{7}{2}$ possibilities are obvious, e.g., the ones with the common leg. In any case, the proof is of the kind shown in the given representative examples below.

(a) *Common hypotenuse h and inradius i.* We use f for the area, p for the perimeter, and α for the angle where $0 < \alpha < \pi/4$. In what follows, s stands for $\sin\alpha$ and c for $\cos\alpha$.

We discuss the function

$$i(\alpha) = \frac{2f}{p} = \frac{2h^2sc}{h + hs + hc}$$

with a fixed value h.

$$\frac{1}{h}\cdot\frac{di}{d\alpha} = \frac{d}{d\alpha}\frac{sc}{1+s+c} = \frac{c^2 + c^3 - s^2 - s^3}{(1+s+c)^2}$$

$$= \frac{(c-s)(c^2+cs+s^2+c+s)}{(1+s+c)^2}$$

is clearly $> 0\ \forall\alpha \in (0, \pi/4)$.

(b) *Common leg-difference $d = |x - y|$ and inradius i.* We consider $\delta := x - y$ to be fixed, then

$$i(x) = \frac{2x(x+\delta)}{2x + \delta + \sqrt{x^2 + (x+\delta)^2}}.$$

With $h = \sqrt{x^2 + (x+\delta)^2}$ and $D = 2x + \delta + h$,

$$D^2 \cdot \frac{di}{dx} = [2(x+\delta) + 2x][2x + \delta + h] - \left[2 + \frac{2x+\delta}{h}\right]\cdot 2x(x+\delta).$$

With $y = x + \delta$ we have

$$D^2 \cdot \frac{di}{dx} = (2x + 2y)(x + y + h) - \left(2 + \frac{x+y}{h}\right)\cdot 2xy$$

$$= 2h^3 + 2(x+y)(h^2 - xy) > 0$$

because $h^2 - xy > 0$ for each right triangle.

Acknowledgements. I would like to thank my students Rudolf Eicher and Ramon Furer for their computer work as well as for their experimental approach to number theory. Both stimulated me in this work. Thanks also to L. A. Szèkeley in Budapest for a private communication. Finally, I would like to mention Prof. Robert Prielipp at the University of Wisconsin Oshkosh, Dr. Baoswan Wong-Dzung in Untersiggenthal, Switzerland and Deynise Griggs in Columbia, South Carolina, who edited my text.

REFERENCES

1. W. A. Adams and L. J. Goldstein, Introduction to Number Theory, Prentice Hall, New Jersey, 1976.
2. A. Anema, Pythagorean triangles with equal perimeters, *Scripta Mathematica*, XV (1949) 89.
3. L. E. Dickson, History of the Theory of Numbers, Volume II, Chelsea Publishing Company, New York, 1971.
4. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford at the Clarendon Press, 1959.
5. L. K. Hua, Introduction to Number Theory, Springer, 1982.
6. N. Koblitz, Introduction to Elliptic Curves and Modular Forms, Springer, New York, Berlin, 1984.
7. A. A. Krishnaswami, On isoperimetrical Pythagorean triangles, *Tôhoku Math. J.*, 27 (1926) 332–348.
8. J. Lambek and L. Moser, On the distribution of Pythagorean triangles, *Pacific J. Math.*, 5 (1955) 73–83.
9. T. Nagell, Introduction to Number Theory, Chelsea Publishing Company, New York, 1964.
10. W. Narkiewic, Number Theory, World Scientific, Singapore, 1977.
11. I. Niven, An Introduction to the Theory of Numbers, Wiley, New York, 1980.
12. P. Théophile Pepin, Solution de quelques problèmes numériques, Memoire della Pontificia Accademica dei Nuovi Lincei, volume settimo, 1891.
13. H. Riesel, Prime Numbers and Computer Methods for Factorization, Birkhäuser, Boston, Basel, Progress in Mathematics, Vol. 57, 1985.
14. A. Robert, Sans tambour ni trompette, *Journal IREM Lyon*, 32 (September 1986).
15. J. Rohlfs, Mathematische Miniaturen 1, Lebendige Zahlen, Über die Summe von zwei Quadraten, Birkhäuser, Boston, Basel, 1981.
16. W. Sierpiński, Pythagorean Triangles, Yeshiva University, New York, 1962.
17. B. M. Stewart, Theory of Numbers, Macmillan, New York, 1964.
18. A. Weil, Number Theory, An Approach Through History, Birkhäuser, Boston, Basel, 1983.
19. W. P. Whitlock Jr., Rational right triangles with equal area, *Scripta Mathematica*, IX (1943), 155–161.
20. _____, Pythagorean triangles with a given difference or sum of sides, *Scripta Mathematica*, XI (1945) 75–80.

A Very Elementary Proof of the Malgrange-Ehrenpreis Theorem

JEAN-PIERRE ROSAY, *University of Wisconsin, Madison*

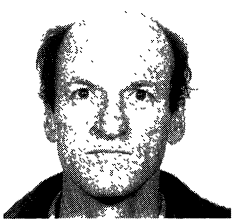
JEAN-PIERRE ROSAY: Ph.D. in 1970 at Grenoble (France), under the supervision of Alain Bernard, in functions algebras.

Postdoctoral positions: University of Kentucky-Lexington, Université Paris-Sud Orsay.

Professor: at Marseilles 1973–86, at the University of Wisconsin-Madison since 1986.

Places visited for a semester: UCLA, University of North Carolina at Chapel-Hill.

Field: Several complex variables.



The aim of this note is to present an easy, self-contained, and totally elementary proof of the fundamental theorem of Malgrange and Ehrenpreis ([1], [6]), on the existence of fundamental solutions for linear partial differential operators with constant coefficients.

By totally elementary the following is meant:

- we will not use any Fourier transform argument
- only few integrations by parts will be used, together with the Hilbert space structure of L^2, and the Riesz representation theorem. Convolutions are used, twice, in a very simple way.
- Basically no knowledge of distribution theory is needed. We first show that for every $g \in L^2_{\mathrm{loc}}(\mathbf{R}^n)$ one can solve the equation $P(D)u = g$, with $u \in L^2_{\mathrm{loc}}(\mathbf{R}^n)$, if $P(D)$ is a (non-zero) linear differential operator on $\mathbf{R}^n$ with constant coefficients.

This is only to say ("by integration by parts") that for all $\varphi \in \mathscr{C}_0^\infty(\mathbf{R}^n)$: $\langle u, P^*(D)\varphi\rangle = \langle g, \varphi\rangle$, where,

$$P(D) = \sum a_J \frac{\partial^{|J|}}{\partial x^J},$$

$$P^*(D) = \sum (-1)^{|J|} \overline{a_J} \frac{\partial^{|J|}}{\partial x^J},$$

$$J = (j_1, \ldots, j_n), \qquad |J| = j_1 + \cdots + j_n,$$

$$\frac{\partial^{|J|}}{\partial x^J} = \frac{\partial^{j_1 + \cdots + j_n}}{\partial x_1^{j_1} \cdots \partial x_n^{j_n}}.$$

Also $\langle\ ,\ \rangle$ denotes the usual scalar product in $L^2(\mathbf{R}^n)$, $\langle u, v\rangle = \int_{\mathbf{R}^n} u\bar{v}$.

- Only the last step, which uses a trivial and well known trick (whose origin, I do not know) requires one to know the definition of a distribution.

I wish to make clear that there is but little novelty in this proof. The basic and almost unique ingredient is Hörmander's inequality ([3] 2.6) (whose proof is so elementary and such a masterpiece). There is an allusion in [8] to an "immediate" proof of the existence of fundamental solutions, using Hörmander's inequality. Here we give the full detail in a way which seems somewhat different. Sections 1 and 2 are taken from Malgrange [7]. The only possibly new point is that we prove

the approximation property in §3, using the same tools, whereas it has usually been done using Fourier transform arguments and using polynomial-exponential solutions ([4] 3.4, [5]). It may be argued that all this was implicit in the literature. Let us then make it explicit!

We do not assume any knowledge of the theory of PDE. But, for instance, it should be noticed that the results on supports in §2 follow immediately from Hölmgren's uniqueness theorem (cf, e.g. [4] 5.3). In fact, they could also be obtained immediately from the uniqueness of solutions of O.D.E., by partial Fourier transform, since only the case of functions with compact support is under consideration. A relatively simple proof of the Malgrange Ehrenpreis theorem, using Fourier transforms, can be found in [2].

Notations. If Ω is an open set in $\mathbf{R}^n$, $\mathscr{C}_0^\infty(\Omega)$ will denote the space of $\mathscr{C}^\infty$ functions with compact support in Ω. The norm and the scalar product in $L^2(\Omega)$ will be respectively denoted by $\| \ \|$ and $\langle \ , \ \rangle$, or $\| \ \|_\Omega$ and $\langle \ , \ \rangle_\Omega$ if some ambiguity is possible.

The adjoint of an operator $P(D)$ is defined as above and denoted by $P^*(D)$.

1. L^2 solvability on bounded open sets in $\mathbf{R}^n$.

1.1. Hörmander's inequality. Let $P(D)$ be a (non-zero) linear differential operator with constant coefficients, of order (or degree) m.

THEOREM. *For every bounded open set Ω in $\mathbf{R}^n$, there exists a constant $C > 0$, such that for every $\varphi \in \mathscr{C}_0^\infty(\Omega)$:*

$$\|P(D)\varphi\| \geqslant C\|\varphi\|.$$

As constant C, one can take $C = |P|_m K_{m,\Omega}$, where

$$|P|_m = \max\{|a_J|, |J| = m\} \left(P(D) = \sum a_J \frac{\partial^{|J|}}{\partial x_J} \right),$$

and $K_{m,\Omega}$ depends only on m (the order of $P(D)$) and the diameter of Ω.

Remark. The proof will use a trick that it is nice to look at first in the case $n = 1$, $\Omega = (0,1)$ and $P(D) = d/dx$! We have to show that $\|\varphi'\| \geqslant C\|\varphi\|$ for all $\varphi \in \mathscr{C}_0^\infty((0,1))$, for some positive constant C.

One has $\langle (x\varphi)', \varphi \rangle = \langle x\varphi', \varphi \rangle + \langle \varphi, \varphi \rangle$. Further, $\langle (x\varphi)', \varphi \rangle = -\langle x\varphi, \varphi' \rangle$, by integration by parts. Hence: $\langle \varphi, \varphi \rangle = -\langle x\varphi', \varphi \rangle - \langle x\varphi, \varphi' \rangle$. Since $|x| < 1$, one gets $\|\varphi\|^2 \leqslant 2\|\varphi'\|\|\varphi\|$, by the Schwarz inequality. So $\|\varphi'\| \geqslant \frac{1}{2}\|\varphi\|$.

The general proof goes by induction on the degree of P.

Proof. If $P(D)$ is a differential operator of order $m \geqslant 0$ on $\mathbf{R}^n$, then for $j = 1, \ldots, n$ define $P_j(D)$ by

$$P(D)(x_j\varphi) = x_j P(D)\varphi + P_j(D)\varphi.$$

The operator $P_j(D)$ is zero if and only if $P(D)$ does not involve any differentiation with respect to x_j, and if non-zero, $P_j(D)$ is of order $< m$. Let $A = \sup_\Omega |x|$. By induction on m, we will show that for every $\varphi \in \mathscr{C}_0^\infty(\Omega)$

$$\|P_j(D)\varphi\| \leqslant 2mA\|P(D)\varphi\|. \qquad (**)$$

Let us first notice that the inequality $(**)$ and the definition of P_j yield: $\|P(D)(x_j\varphi)\| \leqslant (2m+1)A\|P(D)\varphi\|$. Since differential operators with constant coefficients commute, we have, for all $\varphi \in \mathscr{C}_0^\infty(\Omega)$

$$\|P(D)\varphi\|^2 = \langle P(D)\varphi, P(D)\varphi\rangle = \langle \varphi, P^*(D)P(D)\varphi\rangle = \langle \varphi, P(D)P^*(D)\varphi\rangle$$
$$= \langle P^*(D)\varphi, P^*(D)\varphi = \|P^*(D)\varphi\|^2.$$

This is of course just folklore from the theory of normal operators!

The inequality $(**)$ is trivial for $m = 0$, since then $P_j(D) = 0$. Let us assume the inequality $(**)$ has been verified for operators of order at most $m-1$, and let $P(D)$ be an operator of order m. Compute in two different ways $\langle P(D)(x_j\varphi), P_j(D)\varphi\rangle$. From the definition of $P_j(D)$

$$\langle P(D)(x_j\varphi), P_j(D)\varphi\rangle = \langle x_j P(D)\varphi, P_j(D)\varphi\rangle + \|P_j(D)\varphi\|^2.$$

By integration by parts (i.e., the definition of the adjoint) and using the commutation of $P^*(D)$ and $P_j(D)$

$$\langle P(D)(x_j\varphi), P_j(D)\varphi\rangle = \langle P_j^*(D)(x_j\varphi), P^*(D)\varphi\rangle.$$

Therefore,

$$\|P_j(D)\varphi\|^2 = \langle P_j^*(D)(x_j\varphi), P^*(D)\varphi\rangle - \langle x_j P(D)\varphi, P_j(D)\varphi\rangle.$$

As noticed above, the induction hypothesis yields

$$\|P_j^*(D)(x_j\varphi)\| \leqslant (2m-1)A\|P_j(D)\varphi\|.$$

Since

$$|\langle x_j P(D)\varphi, P_j(D)\varphi\rangle| \leqslant A\|P(D)\varphi\|\|P_j(D)\varphi\|,$$

one gets

$$\|P_j(D)\varphi\|^2 \leqslant 2mA\|P_j(D)\varphi\|\|P(D)\varphi\|$$

and this establishes $(**)$.

If $P(D)$ is an operator of order $m \geqslant 1$, there exists $j \in \{1, \ldots, n\}$ so that $P_j(D)$ is of order $m-1$, and $|P_j|_{m-1} \geqslant |P|_m$. The theorem follows then immediately by induction on m.

1.2. L^2 solvability

COROLLARY. *If Ω is a bounded set in $\mathbf{R}^n$, then for every $g \in L^2(\Omega)$ there exists $u \in L^2(\Omega)$ so that $P(D)u = g$.*

Proof. This follows immediately from the inequality on the adjoint $\|P^*(D)\varphi\| \geqslant C\|\varphi\|$, $\varphi \in \mathscr{C}_0^\infty(\Omega)$. Indeed $P(D)u = g$ means that for all $\varphi \in \mathscr{C}_0^\infty(\Omega)$

$$\langle g, \varphi\rangle = \langle u, P^*(D)\varphi\rangle. \qquad (*)$$

Let $E = \{\Psi \in \mathscr{C}_0^\infty(\Omega), \Psi = P^*(D)\varphi \text{ for some } \varphi \in \mathscr{C}_0^\infty(\Omega)\}$. Hörmander's inequality shows that the antilinear form $\Psi \mapsto \langle g, \varphi\rangle$ is well defined and continuous in the L^2 norm. It can therefore be extended to $\bar{E}$, the closure of E in $L^2(\Omega)$. Then the Riesz representation theorem gives the existence of $u \in \bar{E}$ so that $(*)$ holds.

2. Hörmander's inequality with weight and support.

THEOREM. *$P(D)$ and Ω (bounded) are given as previously. There exists $C' > 0$ so that for all $\eta \in R$ and $\varphi \in \mathscr{C}_0^\infty(\Omega)$*

$$\int_\Omega e^{\eta x_1}|P(D)\varphi|^2 \geqslant C'\int_\Omega e^{\eta x_1}|\varphi|^2.$$

Note that C' does not depend on η.

Proof. Apply Hörmander's inequality to $\Psi = e^{(\eta/2)x_1}\varphi$ and the operator $Q(D)$ defined by

$$Q(D)(\Psi) = e^{(\eta/2)x_1}Q(D)[e^{-(\eta/2)x_1}\Psi],$$

which is indeed a constant coefficient operator with the same terms of higher degree as $P(D)$.

COROLLARY 1. *If $\varphi \in \mathscr{C}_0^\infty(\mathbf{R}^n)$ and $P(D)\varphi = 0$ in the half space $\{x_1 > 0\}$, then $\varphi = 0$ in this half space.*

Proof. Take Ω an open set containing the support of φ, and let η tend to $+\infty$ in the inequality.

COROLLARY 2. *Let $\varphi \in \mathscr{C}_0^\infty(\mathbf{R}^n)$, or more generally $\varphi \in L^2(\mathbf{R}^n)$ and assume that φ has compact support. If $P(D)\varphi$ is supported by the ball of radius r around 0, then so is φ.*

In the application $P(D)\varphi$ is given by a function in $L^2(\mathbf{R}^n)$.

Proof. By using translations and rotations, the smooth case follows immediately from Corollary 1, by writing the ball as an intersection of half spaces. To treat the non-smooth case, consider $\chi \in \mathscr{C}_0^\infty(\mathbf{R}^n)$, χ supported by the ball of radius 1 around the origin, and satisfying $\int \chi = 1$. For $\varepsilon > 0$ set

$$\varphi_\varepsilon = \varphi * \chi_\varepsilon = \frac{1}{\varepsilon^n}\int_{\mathbf{R}^n} \varphi(x - t)\chi\left(\frac{t}{\varepsilon}\right) dt.$$

Then $\varphi_\varepsilon \in \mathscr{C}_0^\infty(\mathbf{R}^n)$, $P(D)\varphi_\varepsilon (= [P(D\varphi]_* \chi_\varepsilon)$ is supported by the ball of radius $r + \varepsilon$, and $\varphi_\varepsilon \to \varphi$ in L^2 as $\varepsilon \to 0$. This reduces the problem to the smooth case.

Now that we are at the end of section 2, let us remind the reader that sections 1 and 2 are taken from Malgrange [7].

3. Approximation Theorem. For $\rho > 0$, let B_ρ be the ball of radius ρ around 0 in $\mathbf{R}^n$.

THEOREM. *Let $0 < r < r' < R$. If $v \in L^2(B_{r'})$ and satisfies $P(D)v = 0$ on $B_{r'}$, there exists (v_j), a sequence in $L^2(B_R)$ so that $P(D)v_j = 0$ on B_R and (v_j) tends to v in $L^2(B_r)$ as $j \to \infty$.*

Proof. By smoothing by convolution we can assume v to be smooth, possibly shrinking r' slightly (to preserve $P(D)v = 0$). We have to show that if $g \in L^2(B_R)$ and satisfies $\langle \alpha, g\rangle_{B_r} = 0$ for all $\alpha \in L^2(B_R)$ such that $P(D)\alpha = 0$, then $\langle v, g\rangle_{B_r} = 0$.

Claim. There exists $w \in L^2(B_R)$ so that for all $\varphi \in \mathscr{C}_0^\infty(\mathbf{R}^n)\langle \varphi, g\rangle_{B_r} = \langle P(D)\varphi, w\rangle_{B_R}$.

Proof of the claim. This is just to say that for some constant C

$$|\langle \varphi, g\rangle_{B_r}| \leqslant C\|P(D)\varphi\|_{B_R}.$$

Notice that if $P(D)\varphi = 0$, we indeed have $\langle \varphi, g\rangle = 0$. If $P(D)\varphi \neq 0$, by §1 we can find $\Psi \in L^2(B_R)$ so that $P(D)\Psi = P(D)\varphi$ and $\|\Psi\|_{B_R} \leqslant C_1\|P(D)\varphi\|_{B_R}$ (C_1 a constant). Then $\langle \varphi, g\rangle_{B_r} = \langle \varphi - \Psi, g\rangle_{B_r} + \langle \Psi, g\rangle_{B_r} = \langle \Psi, g\rangle_{B_r}$. Hence $|\langle \varphi, g\rangle_{B_r}| \leqslant C\|P(D)\varphi\|_{B_R}$ with $C = C_1\|g\|$. And the claim is proved.

Pick w as given by the claim. Extend g and w on $\mathbf{R}^n$ to $\tilde{g}$ and $\tilde{w}$, by setting $\tilde{g} = 0$ on $R^n - B_r$ and $\tilde{w} = 0$ on $\mathbf{R}^n - B_R$. We then have (by definition) $\tilde{g} = P^*(D)\tilde{w}$. Since $\tilde{w}$ has compact support, and $P^*(D)\tilde{w}$ is supported by the ball B_r, we conclude from §2 that $w = 0$ on $B_R - B_r$.

Then, take v as in the beginning of the proof, and extend it to be a smooth function on $\mathbf{R}^n$ (but no longer satisfying $P(D)v = 0$ off B_r). One has

$$\langle v, g\rangle_{B_r} = \langle P(D)v, w\rangle_{B_R} = \langle P(D)v, w\rangle_{B_r} = 0. \qquad \text{Q.E.D}$$

4. Solvability of $P(D)u = g$ in $L^2_{\text{loc}}(\mathbf{R}^n)$. The following theorem results immediately from §1 and §3 by a standard procedure (Mittag-Leffler):

THEOREM. *Let $P(D)$ be a (nonzero) constant coefficient linear differential operator on $\mathbf{R}^n$. Then for every $g \in L^2_{\text{loc}}(\mathbf{R}^n)$ there exists $u \in L^2_{\text{loc}}(\mathbf{R}^n)$ such that $P(D)u = g$.*

Proof. As previously, B_p denotes the ball of radius p around 0 in $\mathbf{R}^n$. By §1 there exists $u_1 \in L^2(B_2)$ so that $P(D)u_1 = g$ on B_2. Then, inductively, assuming that u_p has been chosen in $L^2(B_{p+1})$ so that $P(D)u_p = g$, one chooses u_{p+1} in $L^2(B_{p+2})$ in the following way. Let w be an arbitrary solution of $P(D)w = g$, in $L^2(B_{p+2})$. On B_{p+1} one has $P(D)(u_p - w) = 0$. By §3 there exists $v \in L^2(B_{p+2})$ such that: $P(D)v = 0$, and $\|v - (u_p - w)\|B_p \leqslant 1/2^p$. Set $u_{p+1} = v + w$. Then $P(D)u_{p+1} = g$ on B_{p+2}, and $\|u_{p+1} - u_p\|_{B_p} \leqslant 1/2^p$. The sequence (u_p) is obviously convergent in $L^2_{\text{loc}}(\mathbf{R}^n)$, and its limit satisfies $P(D)u = g$.

5. Fundamental solution.

THEOREM (Malgrange-Ehrenpreis [1], [6]). *Every nonzero linear differential operator with constant coefficients on $\mathbf{R}^n$ has a fundamental solution (i.e., a distribution E such that $P(D)E = \delta_0$, δ_0 the Dirac mass at 0).*

At last, we have to assume that the reader knows the definition of a distribution, and the definition of the derivatives of a distribution.

Proof. Let H be the function (product of Heaviside functions) defined on $\mathbf{R}^n$ by: $H(x_1, \ldots, x_n) = 1$ if all the x_j are positive and 0 otherwise. Then $(\partial^n/\partial x_1 \cdots \partial x_n)H = \delta_0$. By §4 there exists $u \in L^2_{\text{loc}}(\mathbf{R}^n)$ so that $P(D)u = H$. Set $E = \partial^n u/(\partial x_1 \cdots \partial x_n)$ (a distribution). Then

$$P(D)E = P(D)\frac{\partial^n u}{\partial x_1 \cdots \partial x_n} = \frac{\partial^n}{\partial x_1 \cdots \partial x_n}(P(D)u) = \delta_0.$$

This ends the proof!

Remark. In the proof one can replace the operator $\partial^n/(\partial x_1 \cdots \partial x_n)$ by any operator which has a fundamental solution that belongs to $L^2_{\text{loc}}(\mathbf{R}^n)$, for example: The Laplacian Δ, if $n = 1, 2, 3$, or $(I - \Delta)^p$ for general n, if $p > n/4$.

REFERENCES

1. L. Ehrenpreis, Solutions of some problems of division I, *Am. J. Math.*, 76 (1954) 883–903.
2. G. B. Folland, Lectures on partial differential equations, *Tata. Inst.*, 1983.
3. L. Hörmander, On the theory of general partial differential operators, *Acta Math.*, 94 (1955) 161–248.
4. _____, Linear partial differential operators, *Grund. Math. Wis.*, 116 (1964) Springer-Verlag.
5. _____, The analysis of linear partial differential operators II, *Grund. Math. Wis.*, 257 (1983) Springer-Verlag.
6. B. Malgrange, Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution, *Ann. Inst. Fourier*, 6 (1955) 271–355.
7. _____, Equations aux dérivées partielles (conférences aux carrés, rédigées par A. Cerezo), Ecole Normale Supérieure, Paris, 1966.
8. F. Treves, Thèse d' Hörmander I, *Séminaire Bourbaki*, 130 (1956).

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

On the Zeros of $\tan^{(k)} z$

LI-CHIEN SHEN
Department of Mathematics, University of Florida, Gainesville, FL 32611

The zeros of the successive derivatives of a given meromorphic function have been investigated by many, notably by G. Pólya, in the past. In particular, from a well-known result of Pólya [1], one can deduce that the set of all the accumulation points of the zeros of $\tan^{(k)} z$, $k = 1, 2, 3, \ldots$, are precisely the vertical lines

$$\{n\pi + iy : -\infty \leqq y \leqq \infty,\ n = 0, \pm 1, \pm 2, \pm 3, \ldots .\}.$$

However, very little is known about the precise nature of those zeros. The primary purpose of this paper is to study the "large" zeros of $\tan^{(k)} z$. Combining the above mentioned result of Pólya and the main theorem of this note one sees that, $\tan^{(k)} z$ has two kinds of zeros; that is, one type accumulates to form a dense set on those vertical lines and the other type diverges at a certain rate (see Theorem 1) to infinity.

It would be interesting to see whether the behavior of the zeros of the successive derivatives of $\tan z$ can be generalized to a larger class of meromorphic functions which have only real poles. Finally, it is worthwhile to point out that $\tan^{(k)} z$ can be written as a special case of the generalized Riemann zeta function (see (6)) which has been studied extensively by many mathematicians.

Since $\tan z$ is a periodic function of period π, we will restrict our study to the vertical strip $S = \{z : -\pi/2 < \operatorname{Re} z \leqq \pi/2\}$. We will use I to denote the imaginary axis and I_0 to denote the line segment $\{z : z = iy, -1 < y < 1\}$, and S^+ and I^+ will be the portions of S and I belonging to the upper half plane, respectively.

It is known that all the zeros of the kth derivative, $\tan^{(k)} z$, $k = 0, 1, 2, 3, \ldots$, $z \in S$, lie on I. To see this, we first observe that, for z in S,

$$\tan z \in I_0 \text{ if and only if } z \in I, \tag{1}$$

and

$$\tan z \text{ maps } I \text{ one to one onto } I_0. \tag{2}$$

We now define a sequence of polynomials as follows:

$$P_1(x) = x;$$

$$P_{k+1}(x) = (1 + x^2) P_k'(x), \qquad k = 1, 2, 3, \ldots .$$

Then

$$\tan^{(k)} z = P_{k+1}(\tan z). \tag{3}$$

It is easy to verify that all the zeros of $P_k(x)$ are simple and lie on the interval $\{z : z = iy, -1 \leqq y \leqq 1\}$. From (1), (2) and (3), we see that for $k \geq 1$ all the zeros of $\tan^{(k)} z$ are simple and belong to I. We further conclude, from (3), that it has $k - 1$ zeros and they are symmetric with respect to the real axis; if k is even, one of its zeros is $z = 0$. So there are exactly $[(k - 1)/2]$ zeros on I^+.

We now order the zeros

$$\left\{z_{k,m} : z_{k,m} = iy_{k,m}, m = 1, 2, \ldots, \left[\frac{k-1}{2}\right]\right\}$$

of $\tan^{(k)} z$ on I^+ so that

$$y_{k,1} > y_{k,2} > y_{k,3} > \cdots .$$

We will prove

THEOREM 1. $z_{k,1} - ik \ln \sqrt{2} \to 0$ *as* $k \to \infty$.

We will first establish some crucial properties of $\tan z$.

Let $s = e^{2iz}$. Then

$$\tan z = \frac{\sin z}{\cos z} = -i\frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}} = -i\left(1 - \frac{2}{1+s}\right). \tag{4}$$

Using the chain rule, $d/dz = 2is\, d/ds$, we see that

$$\tan^{(k)} z = (2i)^{k+1} Q_k(s)/(1+s)^{k+1}, \tag{5}$$

where Q_k is a polynomial of degree k. Since $\tan^{(k)} z$ has only imaginary zeros and $s = e^{2iz}$, we deduce that, with the exception of the zero at $s = 0$, the zeros of Q_k are simple and positive and they are symmetric with respect to the circle $|s| = 1$. That is, if s_0 is a zero of Q_k, then $1/s_0$ is also a zero. Thus it is sufficient to study the zeros of Q_k in the disc $|s| < 1$. From (4),

$$\tan z = -i\left(1 - 2\sum_0^\infty (-1)^n s^n\right), \qquad |s| < 1.$$

Using the chain rule, we obtain

$$\tan^{(k)} z = (2i)^{k+1} f_k(s) \qquad (k \geqq 1), \tag{6}$$

where

$$f_k(s) = \sum_1^\infty (-1)^n n^k s^n = Q_k(s)/(1+s)^{k+1}.$$

We now order the zeros $\{s_{k,m}\}$ of $f_k(s)$ in the unit disc so that

$$0 = s_{k,0} < s_{k,1} < s_{k,2} < \cdots .$$

Then the relation between the zeros of $\tan^{(k)} z$ and $f_k(s)$ is

$$s_{k,m} = \exp(2iz_{k,m}) \qquad m = 1, 2, \ldots . \tag{7}$$

To study the zeros of $f_k(s)$, we first establish the following

LEMMA. *If* $|s| \leqq B^{-k}, 6^{1/3} < B < 2$, *then*

$$|f_k(s) - (-s + 2^k s^2)| < 3^k K_B |s|^3,$$

where K_B *is a constant independent of* k.

we conclude from (10) that

$$|g_k(u) - (-1)^m(u^m - u^{m-1})| \to 0 \text{ as } k \to \infty \tag{11}$$

uniformly on any compact set. If we order the zeros of g_k so that

$$0 = u_{k,0} < u_{k,1} < u_{k,2} < \cdots,$$

we deduce from (11) that

$$\begin{aligned} u_{k,m-1} &\to 1, \\ u_{k,j} &\to 0 \qquad 1 \leqq j \leqq m-2. \end{aligned} \tag{12}$$

The following generalization of Theorem 1 follows immediately.

THEOREM 2. $z_{k,m-1} - (ik/2)\ln(m/(m-1)) \to 0$ *as* $k \to \infty$ *for any fixed* m.

REFERENCES

1. G. Pólya, On the zeros of the derivatives of a function and its analytic character, *Bull. Amer. Math. Soc.*, 49 (1943) 178–191.
2. G. Pólya and G. Szego, Problems and Theorems in Analysis I, Springer-Verlag, New York, 1972.

Two Equal Sums of Three Squares with Equal Products

JOHN B. KELLY
Department of Mathematics, Arizona State University, Tempe, AZ 85287

1. Introduction. In this note we produce a general solution for the Diophantine system

$$\left\{\begin{matrix} x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2 \\ x_1x_2x_3 = y_1y_2y_3 \end{matrix}\right\}. \tag{1}$$

This system is mentioned as an unsolved problem in [3] where it is attributed to John Leech who gave the particular solution {3, 8, 26; 2, 13, 24}.

Actually, the problem goes back to 1909, at least. It was posed as a problem in *L'Intermédiare des Mathématiciens* by U. Bini [2]. Partial solutions were offered by Dubouis and Mathieu.

2. A General Solution. E. T. Bell [1] showed that if $x_1, x_2, x_3, y_1, y_2, y_3$ are integers satisfying $x_1x_2x_3 = y_1y_2y_3$, then there exists an integer matrix

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$$

such that

$$\left\{\begin{matrix} x_i = \prod_{j=1}^{3} t_{ij}, & 1 \leq i \leq 3 \\ y_j = \prod_{i=1}^{3} t_{ij}, & 1 \leq j \leq 3 \end{matrix}\right\}. \tag{2}$$

we conclude from (10) that

$$|g_k(u) - (-1)^m(u^m - u^{m-1})| \to 0 \text{ as } k \to \infty \tag{11}$$

uniformly on any compact set. If we order the zeros of g_k so that

$$0 = u_{k,0} < u_{k,1} < u_{k,2} < \cdots,$$

we deduce from (11) that

$$\begin{aligned} u_{k,m-1} &\to 1, \\ u_{k,j} &\to 0 \qquad 1 \leqq j \leqq m-2. \end{aligned} \tag{12}$$

The following generalization of Theorem 1 follows immediately.

THEOREM 2. $z_{k,m-1} - (ik/2)\ln(m/(m-1)) \to 0$ *as* $k \to \infty$ *for any fixed* m.

REFERENCES

1. G. Pólya, On the zeros of the derivatives of a function and its analytic character, *Bull. Amer. Math. Soc.*, 49 (1943) 178–191.
2. G. Pólya and G. Szego, Problems and Theorems in Analysis I, Springer-Verlag, New York, 1972.

Two Equal Sums of Three Squares with Equal Products

JOHN B. KELLY
Department of Mathematics, Arizona State University, Tempe, AZ 85287

1. Introduction. In this note we produce a general solution for the Diophantine system

$$\left\{\begin{matrix} x_1^2 + x_2^2 + x_3^2 = y_1^2 + y_2^2 + y_3^2 \\ x_1x_2x_3 = y_1y_2y_3 \end{matrix}\right\}. \tag{1}$$

This system is mentioned as an unsolved problem in [3] where it is attributed to John Leech who gave the particular solution {3, 8, 26; 2, 13, 24}.

Actually, the problem goes back to 1909, at least. It was posed as a problem in *L'Intermédiare des Mathématiciens* by U. Bini [2]. Partial solutions were offered by Dubouis and Mathieu.

2. A General Solution. E. T. Bell [1] showed that if $x_1, x_2, x_3, y_1, y_2, y_3$ are integers satisfying $x_1x_2x_3 = y_1y_2y_3$, then there exists an integer matrix

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$$

such that

$$\left\{\begin{matrix} x_i = \prod_{j=1}^{3} t_{ij}, & 1 \leq i \leq 3 \\ y_j = \prod_{i=1}^{3} t_{ij}, & 1 \leq j \leq 3 \end{matrix}\right\}. \tag{2}$$

Insertion of (2) in the first equation in (1) yields an equation in t_{11}, t_{22}, t_{33} of the form

$$A_1 t_{11}^2 + A_2 t_{22}^2 + A_3 t_{33}^2 = 0, \tag{3}$$

where

$$\left.\begin{aligned} A_1 &= t_{12}^2 t_{13}^2 - t_{21}^2 t_{31}^2 \\ A_2 &= t_{21}^2 t_{23}^2 - t_{12}^2 t_{32}^2 \\ A_3 &= t_{31}^2 t_{32}^2 - t_{13}^2 t_{23}^2 \end{aligned}\right\}. \tag{4}$$

If $t_{12}, t_{13}, t_{21}, t_{23}, t_{31}, t_{32}$ are assigned arbitrary values, (3) is the equation of a conic in two-dimensional projective space. All rational points can be found if one is known. It can be checked that (t_{23}, t_{31}, t_{12}) is a solution of (3), i.e., if we put $t_{11} = t_{23}, t_{22} = t_{31}, t_{33} = t_{12}$, then (3) is identically satisfied. This corresponds to the trivial solution of (1) in which (y_1, y_2, y_3) is a permutation of (x_1, x_2, x_3). Any line with rational slope through (t_{23}, t_{31}, t_{12}) intersects the conic at another rational point and any rational point may be so obtained. A parametric form for a line with rational slope through (t_{23}, t_{31}, t_{12}) is

$$\left.\begin{aligned} t_{11} &= \alpha_1 t + t_{23} \\ t_{22} &= \alpha_2 t + t_{31} \\ t_{33} &= \alpha_3 t + t_{12} \end{aligned}\right\}. \tag{5}$$

where $\alpha_1, \alpha_2, \alpha_3$ are arbitrary integers. The line given by (5) intersects the conic given by (3) at a point where

$$t = \frac{-2(\alpha_1 A_1 t_{23} + \alpha_2 A_2 t_{31} + \alpha_3 A_3 t_{12})}{A_1 \alpha_1^2 + A_2 \alpha_2^2 + A_3 \alpha_3^2}.$$

Insertion in (5) gives

$$t_{11} = \frac{-\alpha_1^2 A_1 t_{23} + \alpha_2 A_2(\alpha_2 t_{23} - 2\alpha_1 t_{31}) + \alpha_3 A_3(\alpha_3 t_{23} - 2\alpha_1 t_{12})}{A_1 \alpha_1^2 + A_2 \alpha_2^2 + A_3 \alpha_3^2} \tag{6}$$

with similar expressions for t_{22} and t_{33}.

Equation (6), together with (4), gives the general solution of (3) in rationals in terms of the 9 integer parameters $t_{12}, t_{13}, t_{21}, t_{23}, t_{31}, t_{32}, \alpha_1, \alpha_2$ and α_3. If the denominators in (6) are removed then integral solutions, but not necessarily all integral solutions of (3) are obtained. All integral solutions can be put in the form

$$\left.\begin{aligned} t_{11} &= \beta(-\alpha_1^2 A_1 t_{23} + \alpha_2 A_2(\alpha_2 t_{23} - 2\alpha_1 t_{31}) + \alpha_3 A_3(\alpha_3 t_{23} - 2\alpha_1 t_{12})) \\ t_{22} &= \beta(\alpha_1 A_1(\alpha_1 t_{31} - 2\alpha_2 t_{23}) - \alpha_2^2 A_2 t_{31} + \alpha_3 A_3(\alpha_3 t_{31} - 2\alpha_2 t_{12})) \\ t_{33} &= \beta(\alpha_1 A_1(\alpha_1 t_{12} - 2\alpha_3 t_{23}) + \alpha_2 A_2(\alpha_2 t_{12} - 2\alpha_3 t_{31}) - \alpha_3^2 A_3 t_{12}) \end{aligned}\right\}, \tag{7}$$

where β is rational. The general solution of (1) is given by (7) and (2).

It follows easily from (2) and (4) that if $A_1 = A_2 = A_3 = 0$, only trivial solutions of our original system are obtained. Formula (6) is meaningless when $A_1 \alpha_1^2 + A_2 \alpha_2^2 + A_3 \alpha_3^2 = 0$. A separate analysis of this case yields no solutions not already contained in our general solution. Indeed, only the case where there is a line of

solutions is non-trivial, i.e.,

$$A_1\alpha_1 t_{23} + A_2\alpha_2 t_{31} + A_3\alpha_3 t_{12} = 0.$$

Here, the line (5) is part of the conic (3), yet substitution directly in (7) yields

$$\begin{aligned} t_{11} &= \beta\big(t_{23}\big(A_1\alpha_1^2 + A_2\alpha_2^2 + A_3\alpha_3^2\big) - 2\alpha_1\big(A_1\alpha_1 t_{23} + A_2\alpha_2 t_{31} + A_3\alpha_3 t_{12}\big)\big) \\ &= 0 = t_{22} = t_{33} \end{aligned}$$

so that the points of (5) cannot be obtained from (7) in this way. However, unless $t_{23} = 0$, the point $(-t_{23}, t_{31}, t_{12})$ is not on the line (5), though still on our conic (3), and the points of (5) may be obtained from (7) by choosing $(-t_{23}, t_{31}, t_{12})$ as base point with varying α_1, α_2 and α_3. If $t_{23} = 0$, we may, instead, make a similar use of t_{31} or t_{12} unless they, too, are 0, in which case we have $A_1 = A_2 = A_3 = 0$.

3. Partial Solutions. In [2], Dubouis and Mathieu presented (separately) two-parameter solutions of (1). A linear transformation of the parameters takes one solution into the other so that they are essentially equivalent. Dubouis' solution corresponds, in our notation, to the matrix

$$\begin{bmatrix} 1 & 2b-a & a-b \\ a & 1 & 2a-b \\ a+b & b & 1 \end{bmatrix}.$$

This solution can be obtained from the general solution (7) if one chooses $\alpha_1 = \alpha_2 = \alpha_3 = 1$ and

$$\beta = \frac{1}{8(a^2 - ab + b^2)(b-a)(2a-b)(a-2b)}.$$

Zeitlin [4] gave two one-parameter families of integer solutions of (1). These correspond to the matrices

$$T(a) = \begin{bmatrix} 1 & 1 & 2a+1 \\ 1 & 1 & 3a-2 \\ a & 6a-1 & 1 \end{bmatrix} \text{ and } T'(a) = \begin{bmatrix} 1 & 1 & 2a-1 \\ 1 & 1 & 3a+2 \\ a & 6a+1 & 1 \end{bmatrix},$$

respectively. They can be realized from (7) when $\alpha_1 = \alpha_2 = \alpha_3 = 1$ and β has the respective values

$$\frac{1}{12(3a+1)(a-1)^2} \quad \text{and} \quad \frac{1}{4(3a+1)(a-1)(3a-1)}.$$

Zeitlin's solutions cannot be obtained from those of Dubouis and Mathieu by specializing the parameters. In the latter solutions $x_1^2 + x_2^2 + x_3^2$ is always 5 times a square. This is not the case with Zeitlin's solutions.

REFERENCES

1. E. T. Bell, Reciprocal arrays and Diophantine analysis, *Am. J. Math.*, 55 (1933) 50–66.
2. U. Bini, *L'Intermédiare des Math.*, 16 (1909) 41–43.
3. A. Bremmer and R. K. Guy, A dozen difficult Diophantine dilemmas, this MONTHLY 95 (1988) 31–36.
4. D. Zeitlin, Abstracts presented to American Mathematical Society, April 1988, Issue 57, 239, #843-11-50.

A Coloring Problem

IRA M. GESSEL[1]
Department of Mathematics, Brandeis University, Waltham, MA 02254

Introduction. A well-known algorithm for coloring the vertices of a graph is the "greedy algorithm": given a totally ordered set of colors, each vertex of the graph (taken in some order) is colored with the least color not already used to color an adjacent vertex. When applied to a path graph with at least two vertices, the algorithm uses either 2 or 3 colors, depending on the order in which the vertices are colored.

I. Bouwer and Z. Star [**1**] solved the problem of counting the number of vertex orderings for a path of n vertices (out of $n!$ possible vertex orderings) for which the greedy algorithm uses only two colors. They expressed the number of 2-color vertex orderings in terms of the number $O(n)$ of 2-color vertex orderings in which the first vertex to be colored occurs in an odd position. They then found recurrences for $O(2m)$ and $O(2m+1)$ that led to differential equations for the exponential generating functions

$$G(t) = \sum_{m=0}^{\infty} O(2m+1)\frac{t^{2m+1}}{(2m+1)!}$$

and

$$H(t) = \sum_{m=0}^{\infty} O(2m)\frac{t^{2m}}{(2m)!}$$

(with $O(0) = O(1) = 1$) which they solved to obtain

$$\frac{1}{G(t)} = \frac{1}{t}\left(t - t^2 + \frac{2t}{e^{2t}-1}\right) \tag{1}$$

and

$$H(t) = e^{t^2/2}\exp\left(\int_0^t \tau^2 G(\tau)\,d\tau\right). \tag{2}$$

When a simple generating function can be found by solving a differential equation, it can often be found more directly by a combinatorial argument. Although (1) and (2) don't look particularly simple, (1) can be rewritten as

$$\begin{aligned} G(t) &= \frac{\sinh t}{\cosh t + t\sinh t} \\ &= \sum_{m=0}^{\infty}\frac{t^{2m+1}}{(2m+1)!}\Bigg/\left(1 - \sum_{k=1}^{\infty}(2k-1)\frac{t^{2k}}{(2k)!}\right) \end{aligned} \tag{3}$$

[1]Partially supported by NSF grant DMS-8703600.

tions of the previous paragraph imply that a permutation π of $1, 2, \ldots, n$ is a coloring permutation if and only if the following two conditions are satisfied:

(i) Every hook in the hook factorization of π has even length.

(ii) If n is even, then π has an empty tail.

Thus by (5), the number of coloring permutations whose hooks have lengths $2k_1, \ldots, 2k_r$ and whose tail has length $2m + 1$ is

$$\binom{2k_1 + \cdots + 2k_r + 2m + 1}{2k_1, \ldots, 2k_r, 2m + 1}(2k_1 - 1) \cdots (2k_r - 1). \tag{7}$$

As before, multiplying (7) by $t^{2k_1 + \cdots + 2k_r + 2m + 1}/(2k_1 + \cdots + 2k_r + 2m + 1)!$ and summing on $k_1, \ldots, k_r \geqslant 1$, $m \geqslant 0$, $r \geqslant 0$ yields (3), and (4) can be derived similarly.

REFERENCES

1. I. Bouwer and Z. Star, A question of protocol, *Amer. Math. Monthly*, 95 (1988) 118–121.
2. J. Désarménien, Une autre interprétation du nombre des dérangements, Actes 8^e Séminaire Lotharingien, ed. D. Foata, Publ. I.R.M.A. Strasbourg, 1984, 11–16.
3. I. P. Goulden and D. M. Jackson, Combinatorial Enumeration, Wiley, 1983.

Norm Inequalities for the Powers of a Matrix*

MAN KAM KWONG
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439

A. ZETTL
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

The inequality

$$\|Ax\|^2 \leqslant K\|x\|\,\|A^2x\| \tag{1}$$

has been studied extensively by many authors for operators A acting on a Banach or Hilbert space. Of particular interest are the cases when A is the differentiation operator $Ax = x'$ on the space of differentiable functions x defined on $(-\infty, \infty)$ or $(0, \infty)$, or the forward difference operator $A(x_j) = (x_{j+1} - x_j)$ on the space of infinite sequences $x = \{x_j\}$.

The basic questions are as follows:

1. Given an operator A acting on a normed space X, does there exist a constant $K, 0 \leqslant K < \infty$, such that (1) holds for all x in X?

*This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract W-31-109-Eng-38.

FIG. 1. Graph of K vs. b with $h = 2$

tions of the previous paragraph imply that a permutation π of $1, 2, \ldots, n$ is a coloring permutation if and only if the following two conditions are satisfied:

(i) Every hook in the hook factorization of π has even length.

(ii) If n is even, then π has an empty tail.

Thus by (5), the number of coloring permutations whose hooks have lengths $2k_1, \ldots, 2k_r$ and whose tail has length $2m + 1$ is

$$\binom{2k_1 + \cdots + 2k_r + 2m + 1}{2k_1, \ldots, 2k_r, 2m + 1}(2k_1 - 1) \cdots (2k_r - 1). \tag{7}$$

As before, multiplying (7) by $t^{2k_1 + \cdots + 2k_r + 2m + 1}/(2k_1 + \cdots + 2k_r + 2m + 1)!$ and summing on $k_1, \ldots, k_r \geqslant 1$, $m \geqslant 0$, $r \geqslant 0$ yields (3), and (4) can be derived similarly.

REFERENCES

1. I. Bouwer and Z. Star, A question of protocol, *Amer. Math. Monthly*, 95 (1988) 118–121.
2. J. Désarménien, Une autre interprétation du nombre des dérangements, Actes 8^e Séminaire Lotharingien, ed. D. Foata, Publ. I.R.M.A. Strasbourg, 1984, 11–16.
3. I. P. Goulden and D. M. Jackson, Combinatorial Enumeration, Wiley, 1983.

Norm Inequalities for the Powers of a Matrix*

MAN KAM KWONG
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439

A. ZETTL
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

The inequality

$$\|Ax\|^2 \leqslant K\|x\|\,\|A^2x\| \tag{1}$$

has been studied extensively by many authors for operators A acting on a Banach or Hilbert space. Of particular interest are the cases when A is the differentiation operator $Ax = x'$ on the space of differentiable functions x defined on $(-\infty, \infty)$ or $(0, \infty)$, or the forward difference operator $A(x_j) = (x_{j+1} - x_j)$ on the space of infinite sequences $x = \{x_j\}$.

The basic questions are as follows:

1. Given an operator A acting on a normed space X, does there exist a constant $K, 0 \leqslant K < \infty$, such that (1) holds for all x in X?

*This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract W-31-109-Eng-38.

FIG. 1. Graph of K vs. b with $h = 2$

2. If the answer to the first question is yes, what is the best, i.e., the smallest such K?

In the case of a finite-dimensional normed vector space, question 1 was answered by Ljubic [3] who found that (1) holds with a finite K for all x in X if and only

$$\ker(A) = \ker(A^2), \tag{2}$$

where $\ker(L) = \{x : Lx = 0\}$ denotes the kernel of the operator L. In the simple case where A is a 2×2 matrix, (2) is equivalent to the fact that at most one eigenvalue of A is zero, which holds if and only if either the trace or the determinant is not zero.

The reader is referred to the survey paper [1] for further information and references on inequality (1).

This note reports on our effort in finding the best constant $K = K(A)$ for 2×2 real matrices acting on the Euclidean vector space R^2. One of our purposes is to point out that even for 2×2 matrices, there are still non-trivial problems worth studying. The result we obtained is interesting. Indeed, we are rather surprised to find that $K(A)$ is not given by one single formula and that it is constant for certain ranges of the elements of A.

But the more important purpose of the note is to draw attention to our methodology, the use of symbolic manipulation software. Many people fear that too much reliance on the computer may suppress the need to think and reason, leading to a possible suffocation of creativity. We believe, however, that when used with discretion, computer software can enhance rather than hinder our thinking process. The work reported here is a good example.

The problem of determining $K(A)$ for a matrix has been around for some time. Although there is an elementary and straightforward line of attack, the computation anticipated is huge enough to discourage anyone from pursuing that approach. Very little progress had been made before we enlisted the help of MAPLE to successfully solve the problem for 2×2 matrices. The computation needed for the most general case is too much even for the computer equipment available to us.

MAPLE is symbolic manipulation software developed at the Department of Computer Science, University of Waterloo, Ontario, Canada. The version we used is 4.1 (released in May 1987) installed on a SUN Workstation model 3/60. What it can carry out is exact algebraic (as opposed to numerical) computation, involving symbols used as unknowns. Besides ordinary algebraic operations like multiplying polynomials or rational fractions, it can handle trigonometric and other elementary functions, solve algebraic equations, differentiate and integrate given functions, all with amazing speed and accuracy.

MAPLE did not take over the entire reasoning process. We used MAPLE first to investigate special test matrices. Then after successfully determining $K(A)$ for a few matrices, we mechanized the procedure. Thus we were able to efficiently repeat our experiments with a large number of examples. From the output we detected patterns and formulated conjectures which were later confirmed theoretically.

The role played by MAPLE in this work was in extending our computational capability. Equipped with added confidence, we ventured into and explored avenues that previously appeared inaccessible. In the absence of MAPLE the

problem would probably have remained unsolved for a while longer. On the other hand, human intervention is essential in reformulating the problem to be suitable for MAPLE, in interpreting the output, in stating conjectures, and finally in proving those conjectures. Although a symbolic manipulation program frees us from some technical routines, it has its limitations. Enough problems always remain to challenge our ingenuity. An example is the higher dimensional analog of our problem.

More details of the involvement of MAPLE in this work are presented in [2].

We first show that the general case can be reduced to two special ones. These special cases are then solved completely.

Clearly $K(A) = 0$ if and only if $A = 0$. Let

$$Q(A, x) = \frac{\|Ax\|^2}{\|x\| \, \|A^2 x\|} \quad \text{for } x \in R^2,\ A^2 x \neq 0. \tag{3}$$

From (1), we have

$$K(A) = \max\{Q(A, x) : x \in R^2,\ A^2 x \neq 0\}. \tag{4}$$

LEMMA 1. *Assume that A satisfies* (2). *Then for any orthogonal matrix S and any $\alpha \in R$, $\alpha \neq 0$, we have*

$$K(A) = K(\alpha S A S^{-1}). \tag{5}$$

Proof. The invariance with respect to α is clear from the homogeneous nature of (1). The invariance with respect to orthogonal similarity transformations follows directly from the characterization (4) and the norm preserving properties of orthogonal matrices in Euclidean vector spaces.

Although Lemma 1 sounds obvious once it is stated, it has not been noticed by earlier authors. We suspected its validity only after scrutinizing many special cases. Some simple consequences of Lemma 1 are stated in the next lemma.

LEMMA 2.

1. *If A^{-1} exists, then $K(A) = K(A^{-1})$.*
2. *Assume that A satisfies* (2). *Then*

$$K\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = K\left(\begin{pmatrix} a & c \\ b & d \end{pmatrix}\right) = K\left(\begin{pmatrix} d & b \\ c & a \end{pmatrix}\right)$$
$$= K\left(\begin{pmatrix} a & -b \\ -c & d \end{pmatrix}\right) = K\left(\begin{pmatrix} -a & b \\ c & -d \end{pmatrix}\right). \tag{6}$$

Proof. Part 1 is obvious if we rewrite (3) in the following form which is symmetric with respect to A and A^{-1}:

$$Q(A, y) = \frac{\|y\|^2}{\|A^{-1} y\| \, \|Ay\|} \quad \text{for } y = Ax \in R^2, \qquad Ay \neq 0. \tag{7}$$

Part 2 follows by applying Lemma 1 repeatedly with

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{8}$$

A more useful consequence of Lemma 1 is that the most general matrix A can be reduced to one with equal diagonal elements.

THEOREM 1. *Given any matrix A there exists an orthogonal matrix S such that SAS^{-1} has a constant diagonal, i.e., SAS^{-1} has the form $\begin{pmatrix} \lambda & \mu \\ \nu & \lambda \end{pmatrix}$.*

Proof. Rather than using an orthogonal matrix, we can use some constant multiple of the form $T = \begin{pmatrix} 1 & \beta \\ -\beta & 1 \end{pmatrix}$. Note that $T/\sqrt{1+\beta^2}$ is orthogonal. Then

$$TAT^t = \begin{pmatrix} \beta^2 d + \beta(b+c) + a & -b^2 c + \beta(d-a) + 1 \\ -\beta^2 b + \beta(d-a) + c & \beta^2 a - \beta(b+c) + d \end{pmatrix}. \tag{9}$$

Setting the diagonal elements in (9) equal to each other yields a quadratic equation in β whose discriminant D is

$$D = 4\left((b+c)^2 + (a-d)^2\right) \geqslant 0. \tag{10}$$

Hence a real root exists.

Lemmas 1 and 3 reduce the problem of determining $K(A)$ to one of the two cases $A = \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}$ and $A = \begin{pmatrix} 1 & b \\ c & 1 \end{pmatrix}$. Let us take care of the former case first.

THEOREM 2. *For any $b, c \in R$, $b \neq 0$, and $c \neq 0$,*

$$K\left(\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}\right) = \max\left\{\left|\frac{b}{c}\right|, \left|\frac{c}{b}\right|\right\}. \tag{11}$$

Proof. Note that condition (2) is satisfied under the hypotheses of the theorem. For vectors of the form

$$x = \begin{pmatrix} 1 \\ x \end{pmatrix}, \tag{12}$$

let

$$q(x) = Q(A, x) = \frac{(b^2x^2 + c^2)^2}{(b^2c^2)(1+x^2)^2}. \tag{13}$$

If $b^2 = c^2$, then $q(x) = 1$ for all x in R; hence $K(A) = 1$, and the theorem is proved. If $b^2 \neq c^2$, then $x = 0$ is the only real critical point, the other roots of $q'(x) = 0$ being $\pm ic/b$. Since $q(0) = c^2/b^2$ and $\lim_{x\to\infty} q(x) = b^2/c^2$, it follows that

$$K^2(A) = q_{\max} = \max\left\{\frac{b^2}{c^2}, \frac{c^2}{b^2}\right\}, \tag{14}$$

completing the proof of the theorem.

Observe that Theorem 2 shows that every $K \geqslant 1$ is realized as a best constant in inequality (1). That K cannot be less than 1 for any $A \neq 0$ satisfying (2) can be seen as follows: Let $\|A\| = \max\{\|Ax\| : \|x\| = 1\}$ denote the usual operator norm of A induced by the Euclidean vector norm. Then

$$\|Ax\| \leqslant \|A\|\,\|x\| \quad \text{for all } x. \tag{15}$$

In particular, $\|A^2x\| \leqslant \|A\|\,\|Ax\|$. By compactness, there exists an x_0 such that $\|x_0\| = 1$ and $\|Ax_0\| = 1$. If furthermore $A^2x_0 \neq 0$, then

$$K(A) \geqslant \frac{\|Ax_0\|^2}{\|A^2x_0\|} \geqslant \frac{\|Ax_0\|}{\|A\|} = 1. \tag{16}$$

If $A^2x_0 = 0$, using a sequence of vectors approaching x_0 will yield the same conclusion.

The remaining case is more intriguing.

THEOREM 3. *For $b, h \in R$ with $h \geqslant 0$, let*

$$f_h(b) = K\left(\begin{pmatrix} 1 & b \\ b+h & 1 \end{pmatrix}\right). \tag{17}$$

This function is symmetric about the point $b = -h/2$, i.e.,

$$f_h\left(-b - \frac{h}{2}\right) = f_h\left(b - \frac{h}{2}\right). \tag{18}$$

Let

$$r = \left(1 + \frac{h^2}{4}\right)^{1/2}, \qquad b_1 = \frac{2(1-r)}{h}, \qquad b_2 = \frac{2(1+r)}{h}. \tag{19}$$

We have, for $h > 0$, and $b \geqslant -h/2$,

$$f_h(b) = \begin{cases} r & b_1 \leqslant b \leqslant b_2 \\ \dfrac{|1 - b(b+h)|}{1+b^2} & -\dfrac{h}{2} \leqslant b \leqslant b_1 \text{ and } b_2 \leqslant b. \end{cases} \tag{20}$$

Before discussing the proof we illustrate the theorem by sketching the graph of $f_2(b)$:

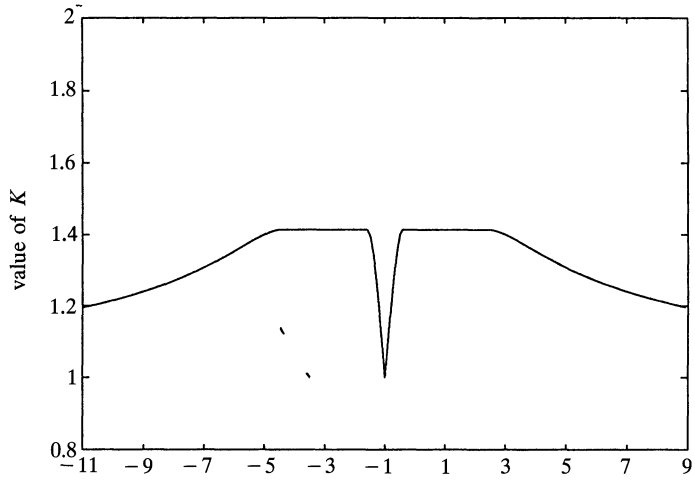


FIG. 1. Graph of K vs. b with $h = 2$.

Proof. We do not have a simple and short proof for formula (20). The only one we have is obtained with the help of MAPLE. Since the computer output is 23 pages, we can give only a brief outline here. We refer the readers to [2] for more details.

We again used normalized vectors of the form (12). The corresponding function to be maximized is then

$$q(x) = \frac{\left((1+bx)^2 + (b+h+x)^2\right)^2}{(1+x^2)(1+b(b+h)+2bx)^2 + (x+(b+h)(2+bx))^2}. \quad (21)$$

All that remains to be done is to obtain the real roots of $q'(x) = 0$, compute the values of $q(x)$ at each of these critical points, as well as at $x = \infty$, and then pick out the largest one. We accomplished these using MAPLE. For $b \neq 0$, $q'(x)$ reduces to a rational function with a polynomial numerator of degree 6. Two of its factors are

$$(bx + 1) \quad \text{and} \quad (x + b + h). \quad (22)$$

When $b = 0$, the numerator of $q'(x)$ is only of degree 5, and the first factor in (22) is absent. Another factor is

$$(1 + b^2)x^2 + (4b + 2h)x + (1 + b^2 + 2bh + h^2), \quad (23)$$

which can be ignored since it has only complex roots. The remaining factor

$$(2b^3h + 3b^2h^2 + 8b^2 + 6bh + bh^3 + h^2)x^2 + (8b^2h + 8bh^2 + 2h^3)x$$
$$+ (2b^3h - 8b^2 + 3b^2h^2 + bh^3 - 10bh - 3h^2) \quad (24)$$

may or may not furnish real roots, depending on the sign of its determinant. More precisely, in the interval $[b_1, b_2]$ and its symmetric partner, the roots are real; otherwise they are complex. In the former case, the maximum of $q(x)$ is attained at one of these two roots. This maximal value, $1 + h^2/4$, is independent of b in the entire interval. It is interesting to note here that MAPLE was unable to simplify the expression for $q(x)$ without external help when one of these roots was substituted for x. In the latter case, the maximum is attained at one of the two roots given by (22).

The fact that the maximum is not attained at $x = \infty$ can be seen as follows. If the maximum is attained at $x = \infty$, then $y = 0$ will be a critical point for the function $q(y)$ obtained by using $x = \binom{y}{1}$ instead of (12). This possibility can be ruled out with the help of MAPLE.

REFERENCES

1. M. K. Kwong and A. Zettl, Landau's inequality for the differential and difference operators, General Inequalities V, Proc. of Conference at Oberwolfach, 1986, Birkhäuser Verlag, Basel, 1987, pp. 367–379.
2. _____, Determining the best constant for the 2×2 matrix Landau inequality using MAPLE, Argonne National Laboratory Preprint #MCS-P7-0988.
3. Ju. I. Ljubic, On inequalities between the powers of a linear operator, *Translations Amer. Math. Soc. Ser. 2*, 40 (1964) 39–84; translated from *Irv. Akad, Nauk. SSSR Ser. Mat.*, 24 (1960) 825–864.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

Numerical Linear Algebra on the HP-28 or How to Lie With Supercalculators

YVES NIEVERGELT
Department of Mathematics, Eastern Washington University, Cheney, WA 99004

1. A sobering experience. The ever easier access to supercalculators enables instructors and students to delegate lengthy calculations to the machines while spending more time thinking about concepts. Yet supercalculators also introduce another phenomenon: large, inconspicuous numerical errors. For example, all digits may be wrong in all the entries of the computed inverse, $A^{\dagger} \approx A^{-1}$, of a nearly singular matrix A, even though $A^{\dagger} \cdot A = I$ to all twelve displayed digits. The analysis of such errors appears in detail in numerical analysis texts [**3**, **14**] and briefly in Hewlett-Packard's manuals [**5**, **6**, **7**], and certain applications do not entail or do not suffer from such errors [**3**, p. 124]. However, most students who use supercalculators never study numerical analysis; consequently, they may need to learn at least that each digital algorithm fails in certain situations, and they should be alerted to the possibility of nonsense in displayed results.

EXAMPLE 1. Consider the following system (abbreviated $A\vec{x} = \vec{b}$ in the sequel):

$$\begin{cases} 888{,}445x + 887{,}112y = 1, \\ 887{,}112x + 885{,}781y = 0. \end{cases}$$

Even with the value of the determinant (one) given to them, no one in a class of twenty juniors and seniors, all majoring in mathematics or computer science, some with supercalculators, produced any values of x and y anywhere close to the solution.

Such a classroom experience demonstrates that, while supercalculators have a key purporting to solve linear systems, students have problems choosing the best numerical method and assessing its accuracy. This phenomenon—the absence of any infallible algorithm—constitutes a new development. For, hitherto, the keys on calculators represented only real, complex, direct, or inverse elementary functions and returned results accurate to all but perhaps the last displayed digit [**5**, p. 179]. Now, some keys perform operations with matrices, and with these, all digits may be wrong.

The next section shows the extent of the problem, and the last section suggests explanations that may help students think about concepts while using supercalculators.

2. Unsettling experiments. Widely different algorithms applied to the system in Example 1 produce wildly different results, as the following experiments show. (Since the last few displayed digits may depend upon the particular implementation of each algorithm, readers may request from the author listings of the programs used below.)

2.1. *Supercalculator keys.* On Hewlett-Packard's HP-28 (C & S) machines, the division key ($\div$) also solves linear systems. For the system in Example 1, it yields

$$x = 1{,}279{,}847.88527 \quad \text{and} \quad y = -1{,}281{,}771.0215.$$

(Then the experienced user may apply iterative refinements [**3**, p. 126; **5**, p. 120; **6**, p. 111; **7**, p. 68], which wander in the range $884{,}000 < x < 887{,}000$ and $-888{,}000 < y < -886{,}000$.)

2.2. *Gaussian elimination.* Programmed on the HP-28, Gaussian elimination and backward substitution gives (without pivoting, without scaling)

$$x = 997{,}001.502623 \quad \text{and} \quad y = -998{,}499.625748.$$

(Hence, iterative refinements behave as described in the preceeding paragraph.)

2.3. *Singular value decomposition.* Also programmed on the HP-28, the singular value decomposition, with annihilation of the smaller singular value [**10**, p. 56–57] produces

$$x = 2.82236169408 \times 10^{-7} \quad \text{and} \quad y = 2.81812709528 \times 10^{-7}$$

This result accurately represents the orthogonal projection of the solution on the eigenspace of A with the larger eigenvalue, $\lambda_{\max} \approx 1{,}774{,}226.00002$, spanned by $\vec{\mathbf{v}}_{\max} \approx (0.707637442, 0.706575721)^T$, and ignores (as it should) the other, numerically unstable projection on the eigenspace with the smaller eigenvalue, $\lambda_{\min} \approx 5.63513515643 \times 10^{-7}$, spanned by $\vec{\mathbf{v}}_{\min} \approx (0.706575721, -0.707637442)^T$. This amounts to one accurate coordinate instead of two inaccurate ones. To check these results, observe that the symmetric matrix A is well conditioned for the eigenvalue problem because it may be diagonalized by an orthonormal change of coordinates [**15**, p. 87]. Also, observe that

$$A \approx \tilde{A} := 887{,}112.5 \times \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

with eigenvalues 1,774,225 and 0, and eigenvectors $(1/\sqrt{2}, 1/\sqrt{2})^T$ and $(1/\sqrt{2}, -1/\sqrt{2})^T$.

2.4. *Eigenvalue diagonalization.* Since the matrix A in Example 1 is symmetric (and positive definite), it has a matrix $\Omega = (\vec{\mathbf{v}}_{\max}, \vec{\mathbf{v}}_{\min})$ of orthonormal eigenvectors. Programming Jacobi's method [**14**, p. 341–344] to compute Ω, and inverting the (positive) diagonal matrix $\Omega^T A \Omega$, all on the HP-28, gives $\vec{\mathbf{x}} = \Omega(\Omega^T A \Omega)^{-1} \Omega^T \vec{\mathbf{b}}$ with

$$x = 885{,}957.898511 \quad \text{and} \quad y = -887{,}289.164320.$$

2.5. *Cramer's rule on the HP-28.* The HP-28 correctly evaluates the determinant (⟨⟨A DET⟩⟩ returns 1) and the transpose (⟨⟨A TRN⟩⟩); hence changing two signs yields A^{-1}, in the first column of which appears, finally, *the solution*:

$$x = 885{,}781 \quad \text{and} \quad y = -887{,}112.$$

2.6. *Question.* What might instructors and textbooks tell students, who carefully copy down their supercalculators' results without ever suspecting that they are copying random digits? The next section suggests answers at various levels.

3. Close encounters of the singular kind. Though the matrix A in Example 1 has only six-digit integral entries, the following phenomenon explains why computations carrying twice as many digits *must* err: there exists a singular matrix S that lies so close to A that the supercalculator cannot distinguish A from S. Thus, computing $\vec{\mathbf{x}} = A^{-1}\vec{\mathbf{b}}$ through the command $bA \div$ amounts to the nonsense $bS \div$ [**5**, p. 102].

EXAMPLE 2. Retaining at most twelve digits from each input, the HP-28 does not distinguish A from the singular matrix

$$S := \begin{pmatrix} 888445 - 1/3548450 & 887112 - 1/3548450 \\ 887112 - 1/3548450 & 885781 - 1/3548450 \end{pmatrix}$$

$$\approx \begin{pmatrix} 888444.9999997 & 887111.9999997 \\ 887111.9999997 & 885780.9999997 \end{pmatrix}.$$

In general, the same phenomenon occurs if the matrix A lies too close to the set of singular matrices, D. To visualize this situation, restrict the discussion to the space $\mathbb{S}(2, \mathbb{R})$ of symmetric 2×2 real matrices, which has dimension three and may be pictured as $\mathbb{R}^3$:

$$\mathbb{S}(2, \mathbb{R}) = \left\{ S \in \mathbb{M}_{2\times 2}(\mathbb{R}) : S = \begin{pmatrix} u & w \\ w & v \end{pmatrix};\ u, v, w \in \mathbb{R} \right\} \cong \mathbb{R}^3.$$

In that space, the singular matrices form the variety D where the determinant $uv - w^2$ vanishes. The change of coordinates $u := r + s$ and $v := r - s$ reveals that D is the right circular cone about the r-axis, with vertex at $\vec{\mathbf{0}}$ and equation $r^2 = s^2 + w^2$.

Thus, computations will *fail* if the distance from the point A to the cone D is less than the relative accuracy of the supercalculator. The next section explains how to calculate this distance from other quantities.

4. Condition numbers. Consider a linear system $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ and denote by $\vec{\mathbf{r}}$ the result returned by a supercalculator attempting to solve for $\vec{\mathbf{x}}$. For convenience, let $\Delta\vec{\mathbf{x}} := \vec{\mathbf{r}} - \vec{\mathbf{x}}$ and $\Delta\vec{\mathbf{b}} := A\vec{\mathbf{r}} - \vec{\mathbf{b}} = A(\vec{\mathbf{x}} + \Delta\vec{\mathbf{x}}) - \vec{\mathbf{b}} = A(\Delta\vec{\mathbf{x}})$. One method to estimate the discrepancy $\Delta\vec{\mathbf{x}}$ between the result $\vec{\mathbf{r}}$ and the solution $\vec{\mathbf{x}}$, widely applied by students and considered as a benchmark by some mathematicians [**4**, p. 1341; **12**, p. 992; **13**, p. 395], consists of substituting $\vec{\mathbf{r}}$ for $\vec{\mathbf{x}}$ and evaluating the "residual" $\Delta\vec{\mathbf{b}} = A\vec{\mathbf{r}} - \vec{\mathbf{b}}$. A standard argument [**14**, p. 175–183] reveals the *worthlessness* of the method, by means of "subordinate norms."

DEFINITION 1. Let $\| \ \|$ represent any norm on $\mathbb{R}^n$. Its **subordinate norm**, also denoted by $\| \ \|$, is defined on the matrices in $\mathbf{M}_{n\times n}(\mathbb{R})$ by

$$\|A\| := \max\{\|A\vec{\mathbf{v}}\| : \vec{\mathbf{v}} \in \mathbb{R}^n,\ \|\vec{\mathbf{v}}\| = 1\} = \max\left\{\|A\vec{\mathbf{v}}\|/\|\vec{\mathbf{v}}\| : \vec{\mathbf{v}} \in \mathbb{R}^n \setminus \{\vec{\mathbf{0}}\}\right\}.$$

A subordinate norm measures the maximum amount by which a matrix may stretch any nonzero vector $\vec{\mathbf{x}}$:

$$\|A\vec{\mathbf{x}}\| = \frac{\|A\vec{\mathbf{x}}\|}{\|\vec{\mathbf{x}}\|}\|\vec{\mathbf{x}}\| \leqslant \max\left\{\|A\vec{\mathbf{v}}\|/\|\vec{\mathbf{v}}\| : \vec{\mathbf{v}} \in \mathbb{R}^n \setminus \{\vec{\mathbf{0}}\}\right\} \cdot \|\vec{\mathbf{x}}\| = \|A\| \cdot \|\vec{\mathbf{x}}\|.$$

In the particular case $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$, this inequality gives $\|\vec{\mathbf{b}}\| = \|A\vec{\mathbf{x}}\| \leqslant \|A\| \cdot \|\vec{\mathbf{x}}\|$. If, moreover, A has an inverse, A^{-1}, then the same inequality applied to $\Delta\vec{\mathbf{x}} =$

Then let $\vec{\mathbf{x}}_0 := A^{-1}\vec{\mathbf{y}}_0$ and choose a matrix $F \in \mathbf{M}_{n\times n}(\mathbb{R})$ such that $F\vec{\mathbf{x}}_0 = A\mathbf{x}_0$ and $\|F\| = 1/\|A^{-1}\|$. (Observe that $\|F\vec{\mathbf{x}}_0\| = \|A\vec{\mathbf{x}}_0\| = \|\vec{\mathbf{y}}_0\| = \|A^{-1}\vec{\mathbf{y}}_0\|/\|A^{-1}\| = \|\vec{\mathbf{x}}_0\|/\|A^{-1}\|$; consequently, $\|F\| = 1/\|A^{-1}\|$ on $\mathrm{Span}(\vec{\mathbf{x}}_0)$. Then construct F by *finitely* many such induction steps as those in the standard proof of the Hahn-Banach Theorem [**1**, p. 149–150].) Finally, let $S := A - F$. Then $S \in D$ because $S\vec{\mathbf{x}}_0 = A\vec{\mathbf{x}}_0 - F\vec{\mathbf{x}}_0 = \vec{\mathbf{0}}$ and $\vec{\mathbf{x}}_0 = A^{-1}\vec{\mathbf{y}}_0 \neq \vec{\mathbf{0}}$. Also, $\|A - S\| = \|F\| = 1/\|A^{-1}\|$. Therefore, $d(A, D) = 1/\|A^{-1}\|$.

EXAMPLE 4. With the "maximum" norm on $\mathbb{R}^2$ and with A as in the previous examples, $d(A, D) = 1/\|A^{-1}\| = 1/1{,}775{,}557$. Incidentally, notice that $1/1{,}775{,}557 < 1/1{,}774{,}225 = 2/3{,}548{,}450 = \|A - S\|$, which means that the singular matrix S in Example 2 is not quite the one closest to A (with distances measured by the row-sum norm).

5. Conclusions for students. Unfortunately, there exists no general algorithm to estimate errors [**9**]. Therefore, the use of a supercalculator, and the scant discussion in its manual about the possible ill-condition of the matrix under consideration, demand an understanding of why and how computations may fail (in other words, theory). At the least, a manual (or an instructor) could recommend the following simple test to reveal the possibility of hidden computational failures: compute the logarithm of the condition number, $\mathrm{Log}_{10}(\mathrm{cond}(A))$. (On the HP-28, run ≪ A RNRM LOG A INV RNRM LOG + ≫ , replacing $\mathrm{Log}_{10}(\|A\| \cdot \|A^{-1}\|)$ by $\mathrm{Log}_{10}\|A\| + \mathrm{Log}_{10}\|A^{-1}\|$ to avoid overflow.) If the result exceeds the number of digits carried by the supercalculator, then the computation of $\vec{\mathbf{x}} = A^{-1}\vec{\mathbf{b}}$ through *bA* ÷ might be meaningless, as explained in Example 3. For instance, for the matrix A in the preceding examples, $\mathrm{Log}_{10}(\mathrm{cond}(A))$ returns 12.6585012324, which exceeds 12. Observe that the same possibility of failure remains even if the large value of the test just described arises from errors in the computation of A^{-1} (because then an example of a failure has already occurred).

REFERENCES

1. Gerald B. Folland, Real Analysis: Modern Techniques and Their Applications, Wiley, New York, 1984.
2. I. C. Gohberg and M. G. Krein, Fundamental aspects of defect numbers, root numbers and indexes of linear operators, *Uspekhi Matematicheskikh Nauk* (N.S.), Vol. 12, No. 2 (March–April 1957), 74 p. 43–118 (Russian). [MR, Vol. 20 (1959), p. 572, #3459.]
3. Gene H. Golub and Charles F. van Loan, Matrix Computations, 2nd ed., John Hopkins University Press, Baltimore, 1989.
4. Eugene A. Herman, Mathematica, a rèview, *Notices of the AMS*, 35 (November 1988) 1334–1344.
5. Hewlett-Packard, HP-15C Advanced Functions Handbook, part no. 00015-90011 Rev B, Hewlett-Packard, Corvallis, OR 1984.
6. _____, HP-28C Reference Manual, part no. 00028-90021, Hewlett-Packard, Corvallis, OR, 1986.
7. _____, HP-28S Advanced Scientific Calculator Reference Manual, part no. 00028-90068, Hewlett-Packard, Corvallis, OR, 1987.
8. William M. Kahan, Numerical Linear Algebra, *Canadian Mathematical Bulletin* (1966) 757–301.
9. _____, The Regrettable Failure of Automated Error Analysis, notes distributed at the Computers & Mathematics Conference at MIT on 13 June 1989.
10. William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1986.
11. Danuta Przeworska-Rolewicz and Stefan Rolewicz, Equations in Linear Spaces, Mathematical Monographs, Vol. 47, PWN-Polish Scientific Publishers, Warszawa, 1968.

12. Barry Simon and Richard M. Wilson, Supercalculators on the PC, *Notices of the AMS*, 35 (September 1988) 978–1001.
13. Raymond F. Smith, Review of MathView Professional, *Notices of the AMS*, 36 (April 1989) 393–399.
14. Josef Stoer and Roland Bulirsch, Introduction to Numerical Analysis, corrected 2nd printing, Springer-Verlag, New York, 1983.
15. James Hardy Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1988.

Gram-Schmidt Orthogonalization by Gauss Elimination

LYLE PURSELL and S. Y. TRIMBLE
Department of Mathematics and Statistics, University of Missouri, Rolla, MO 65401

1. Introduction. In classroom work, attempts by students to use the Gram-Schmidt process typically lead to messy and unorganized computations. Little effort is made to connect this process to the already familiar process of Gauss elimination. In what follows, we show how to use Gauss elimination to orthogonalize a given set of vectors. While it is not numerically preferable to existing algorithms, our algorithm is simple and has some surprisingly happy side effects. These are discussed in the conclusion.

2. An Example. Suppose

$$a_1 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 := \begin{bmatrix} -2 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 5 \end{bmatrix}.$$

Let A be the 4×3 matrix containing a_k as its kth column. Then

$$A^TA = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 14 & 6 \\ 6 & 6 & 28 \end{bmatrix}.$$

Applying Gauss elimination on $[A^TA|A^T]$, we have

$$\left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 4 & 14 & 6 & -2 & 3 & 0 & 1 \\ 6 & 6 & 28 & 1 & 1 & 1 & 5 \end{array}\right] \rightarrow \left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 0 & 6 & -6 & -2 & 1 & 0 & -1 \\ 0 & -6 & 10 & 1 & -2 & 1 & 2 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 0 & 6 & -6 & -2 & 1 & 0 & -1 \\ 0 & 0 & 4 & -1 & -1 & 1 & 1 \end{array}\right].$$

The rows on the right side of this last matrix are mutually orthogonal, and the squares of the lengths of these rows appear on the diagonal of the left side of the matrix. The readers can check for themselves that these rows are precisely the vectors produced by using Gram-Schmidt orthogonalization (without normalizing) on a_1, a_2, and a_3. Note that the only row operation used in our process is the subtraction from one row of a scalar multiple of a row above it. We do not use the interchange of two rows or the multiplication of a row by a scalar.

12. Barry Simon and Richard M. Wilson, Supercalculators on the PC, *Notices of the AMS*, 35 (September 1988) 978–1001.
13. Raymond F. Smith, Review of MathView Professional, *Notices of the AMS*, 36 (April 1989) 393–399.
14. Josef Stoer and Roland Bulirsch, Introduction to Numerical Analysis, corrected 2nd printing, Springer-Verlag, New York, 1983.
15. James Hardy Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1988.

Gram-Schmidt Orthogonalization by Gauss Elimination

LYLE PURSELL and S. Y. TRIMBLE
Department of Mathematics and Statistics, University of Missouri, Rolla, MO 65401

1. Introduction. In classroom work, attempts by students to use the Gram-Schmidt process typically lead to messy and unorganized computations. Little effort is made to connect this process to the already familiar process of Gauss elimination. In what follows, we show how to use Gauss elimination to orthogonalize a given set of vectors. While it is not numerically preferable to existing algorithms, our algorithm is simple and has some surprisingly happy side effects. These are discussed in the conclusion.

2. An Example. Suppose

$$a_1 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 := \begin{bmatrix} -2 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 5 \end{bmatrix}.$$

Let A be the 4×3 matrix containing a_k as its kth column. Then

$$A^TA = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 14 & 6 \\ 6 & 6 & 28 \end{bmatrix}.$$

Applying Gauss elimination on $[A^TA|A^T]$, we have

$$\left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 4 & 14 & 6 & -2 & 3 & 0 & 1 \\ 6 & 6 & 28 & 1 & 1 & 1 & 5 \end{array}\right] \rightarrow \left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 0 & 6 & -6 & -2 & 1 & 0 & -1 \\ 0 & -6 & 10 & 1 & -2 & 1 & 2 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|cccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 0 & 6 & -6 & -2 & 1 & 0 & -1 \\ 0 & 0 & 4 & -1 & -1 & 1 & 1 \end{array}\right].$$

The rows on the right side of this last matrix are mutually orthogonal, and the squares of the lengths of these rows appear on the diagonal of the left side of the matrix. The readers can check for themselves that these rows are precisely the vectors produced by using Gram-Schmidt orthogonalization (without normalizing) on a_1, a_2, and a_3. Note that the only row operation used in our process is the subtraction from one row of a scalar multiple of a row above it. We do not use the interchange of two rows or the multiplication of a row by a scalar.

3. The justification. Let m and n be positive integers, let $\mathbf{R}^m$ be the Euclidean m-space of real vectors, and let $a_1, \ldots, a_n$ be n linearly independent vectors in $\mathbf{R}^m$. Define A to be the $m \times n$ matrix whose jth column is a_j, and let $\mathscr{R}(A)$ be the column space of A.

The linear independence of the columns of A implies that A^TA is positive definite. It follows that Gauss elimination can be carried out on A^TA without row exchanges [**1**, p. 413], [**3**, p. 331]. In fact, the only row operation used is that of subtracting from one row a scalar multiple of a row *above* it. Hence, there are unique $n \times n$ matrices, L and U, such that

$$A^TA = LU, \tag{1}$$

L being a lower triangular matrix with 1's on the diagonal and U being an upper triangular, invertible matrix. Define

$$\hat{Q} := A(L^{-1})^T.$$

Using (1), it follows that $\hat{Q}^T\hat{Q} = U(L^{-1})^T$. Since $\hat{Q}^T\hat{Q}$ is symmetric and $U(L^{-1})^T$ is upper triangular, $\hat{Q}^T\hat{Q}$ must be a diagonal matrix, i.e., the columns of $\hat{Q}$ are orthogonal. Since $\hat{Q}^T = L^{-1}A^T$, it follows that $\hat{Q}^T$ will be computed automatically if Gauss elimination is done on the $n \times (n + m)$ matrix, $[A^TA|A^T]$. Specifically, Gauss elimination turns $[A^TA|A^T]$ into $[U|\hat{Q}^T]$.

Actually, more than this happens. Recall that, since the columns of A are linearly independent, there are two unique matrices, Q and R, such that

$$A = QR, \tag{2}$$

Q being an $m \times n$ matrix with orthonormal columns and R being an upper triangular matrix with positive entries on its diagonal. The columns of Q are precisely the vectors obtained from doing the Gram-Schmidt process on the columns of A [**1**, p. 237], [**3**, p. 175]. Defining

$$\hat{R} := L^T,$$

it follows at once from our definition of $\hat{Q}$ that

$$A = \hat{Q}\hat{R}. \tag{3}$$

Comparing (2) and (3) and defining

$$D := \hat{Q}^T\hat{Q},$$

it is immediate that

$$Q = \hat{Q}D^{-1/2} \quad \text{and} \quad R = D^{1/2}\hat{R}.$$

Since $U = DL^T$, the entries on the diagonals of D and U are the same, i.e., the squares of the lengths of the columns of $\hat{Q}$ are the diagonal entries of U. (See [**3**, p. 196].) Since the entries below the diagonal in L are the negatives of the multipliers used in Gauss elimination as A^TA is turned into U, and since $\hat{R} = L^T$, the $\hat{Q}\hat{R}$ factorization of A is obtained with no extra work. In our earlier example,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}.$$

Hence, as a by-product of our method of doing Gram-Schmidt orthogonalization

on the columns of A, we have

$$A = \begin{bmatrix} 0 & -2 & 1 \\ 1 & 3 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 0 & -2 & -1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Our process can still be used even if the vectors, $a_1, \ldots, a_n$, are not linearly independent. In doing Gauss elimination on a matrix, it is normally necessary to switch rows if 0 should appear on the diagonal in a place where a non-zero pivot is needed. However, as the following theorem shows, if the matrix is positive semidefinite, it is never necessary to make this switch.

THEOREM. *Let B be a positive semidefinite matrix. Then there are a lower triangular, invertible matrix, L, and an upper triangular matrix, U, such that*

$$B = LU. \qquad (*)$$

Further, if the diagonal entry of the ith row of U is 0, then all of the entries in that row of U are 0.

Proof. Suppose 0 appears as the diagonal entry on the kth row of the matrix that is the result of doing Gauss elimination on the first $k-1$ rows of B and that no interchange of rows has yet taken place in doing this Gauss elimination. Then $B = L_k B_k$, where L_k is the identity matrix, except that entries below its diagonal in its first $k-1$ columns need not be 0, and the entries of B_k below its diagonal in its first $k-1$ columns are all 0. (If $k = 1$, then L_k is the identity matrix and $B_k = B$.) Further, the diagonal entry in the kth row of B_k is 0.

Now $(L_k^{-1})^T$ is the identity matrix except that entries to the right of its diagonal in its first $k-1$ rows need not be 0. The definition of matrix multiplication then implies that the kth rows of B_k and $B_k(L_k^{-1})^T$ are the same. In particular, the kth diagonal entry of $B_k(L_k^{-1})^T$ is 0.

Since $B_k(L_k^{-1})^T = L_k^{-1}B(L_k^{-1})^T$, it follows that $B_k(L_k^{-1})^T$ is positive semidefinite. So, there is a matrix, C, such that $B_k(L_k^{-1})^T = C^TC$, [**3**, pp. 339–340]. Let e_k be the column vector with the same length as a column of B, with 1 in its kth position, and with 0's elsewhere. Then the kth diagonal entry of $B_k(L_k^{-1})^T$ is $e_k^T(C^TC)e_k$. From above, this is 0. Hence, $Ce_k = 0$. Since Ce_k is the kth column of C, it follows from the definition of matrix multiplication that the kth row of C^TC is 0. Hence, the same is true of the kth row of B_k.

Since $B_k(L_k^{-1})^T$ is symmetric, its kth column must also be 0. But the definition of matrix multiplication and the shapes of B_k and $(L_k^{-1})^T$ imply that the entries in B_k and $B_k(L_k^{-1})^T$ in their kth columns below the diagonal must be the same, i.e., the entries in B_k in its kth column below its diagonal are all 0. Consequently, there is no need to do Gauss elimination on this column. The Gauss algorithm now proceeds to the $(k+1)$st column with $L_k = L_{k+1}$, $B_k = B_{k+1}$, and $B = L_{k+1}B_{k+1}$. Continuing this process, we conclude that $(*)$ holds and that, if the diagonal entry of the ith row of U is 0, then the ith row itself is 0. □

Now A^TA is always positive semidefinite, so the previous theorem shows that (1) always holds, and the proof shows that (1) can be obtained by Gauss elimination without any row exchanges. The definition of $\hat{Q}$ below (1) does not depend on the existence of U^{-1}, and in all cases the columns of $\hat{Q}$ form an orthogonal set of vectors. The way in which Gauss elimination is done shows that the ith column of

Applying Gauss elimination on A^T, we discover that, if

$$b_1 := \begin{bmatrix} -1 \\ -1 \\ -3 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad b_2 := \begin{bmatrix} 0 \\ 0 \\ 4 \\ 0 \\ 1 \end{bmatrix},$$

then $\{b_1, b_2\}$ is a basis for $\mathcal{N}(A^T)$. Defining

$$B := \begin{bmatrix} 0 & -2 & 1 & -1 & 0 \\ 1 & 3 & 1 & -1 & 0 \\ 0 & 0 & 1 & -3 & 4 \\ 1 & 1 & 5 & 1 & 0 \\ 0 & 0 & -4 & 0 & 1 \end{bmatrix}$$

and using Gauss elimination on $[B^TB|B^T]$, we have

$$\left[\begin{array}{ccccc|ccccc} 2 & 4 & 6 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 4 & 14 & 6 & 0 & 0 & -2 & 3 & 0 & 1 & 0 \\ 6 & 6 & 44 & 0 & 0 & 1 & 1 & 1 & 5 & -4 \\ 0 & 0 & 0 & 12 & -12 & -1 & -1 & -3 & 1 & 0 \\ 0 & 0 & 0 & -12 & 17 & 0 & 0 & 4 & 0 & 1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccccc|ccccc} 2 & 4 & 6 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 6 & -6 & 0 & 0 & -2 & 1 & 0 & -1 & 0 \\ 0 & 0 & 20 & 0 & 0 & -1 & -1 & 1 & 1 & -4 \\ 0 & 0 & 0 & 12 & -12 & -1 & -1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 & -1 & -1 & 1 & 1 & 1 \end{array}\right].$$

So, the desired orthogonal basis is

$$q_1 := \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad q_2 := \frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad q_3 := \frac{1}{\sqrt{20}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ -4 \end{bmatrix},$$

$$q_4 := \frac{1}{\sqrt{12}} \begin{bmatrix} -1 \\ -1 \\ -3 \\ 1 \\ 0 \end{bmatrix}, \quad q_5 := \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

6. Conclusion. We would like to make it clear we are *not* proposing that the current method ([**1**, p. 229], [**3**, p. 172]) of presenting the Gram-Schmidt process be abandoned; students have a real need to understand the geometry of this process. We *are* suggesting that our method, presented above, of hand-calculating the vectors is superior to current hand-calculating methods. Its value does not lie in a shortening of the students' arithmetic work nor in greater computer precision. (Our calculations suggest that the number of multiplications is actually *increased* by $O(n^3/6)$. And while our method is equivalent to the *modified Gram-Schmidt process* ([**1**, p. 233], [**2**], [**3**, p. 181]), the increased precision of that modified process is lost because of the necessity of finding pivots in doing the LU-factorization of A^TA.) Instead, the value of our method lies in the organized format that Gauss elimination naturally brings to a problem.

Perhaps more importantly, by being exposed to our method, the students gain a much deeper understanding and appreciation of the uses of Gauss elimination. Our method unifies the ideas of two apparently unconnected processes and demonstrates the power of viewing matrix transformations as factorizations.

REFERENCES

1. Ben Noble and James W. Daniel, Applied Linear Algebra, 3rd edition, Prentice-Hall, Englewood Cliffs, N.J., 1988.
2. John R. Rice, Experiments on Gram-Schmidt orthogonalization, *Math. Comp.*, 20 (1966) 325–328.
3. Gilbert Strang, Linear Algebra and Its Applications, 3rd edition, Harcourt Brace Jovanovich, San Diego, 1988.

On Finite Groups and Finite Fields

J. D. Reid
Department of Mathematics, Wesleyan University, Middletown, CT 06457

Underlying one of the proofs, perhaps the most elementary one, of the fact that the multiplicative group of nonzero elements in a finite field is cyclic is an idea that we think perhaps deserves to be made more explicit. This can be done in terms of certain functions defined by finite groups, functions that might have some interest in their own right (see, for example, the *Remarks* below). Furthermore this approach to cyclic groups suggests an analogous treatment of a more purely field theoretic problem, namely the derivation of a formula for the number of irreducible polynomials of degree n over a finite field of cardinality q. While all this is essentially just a reformulation of standard results (cf. [**1**], pp. 259–261, or [**3**], p. 4, for example), we hope nonetheless that some readers will find this point of view interesting.

Groups. Suppose then that G is any group of order n. For elements a of G we denote by $0(a)$ the order of a and by $\langle a\rangle$ the cyclic subgroup of G generated by a. Further, let $\#(G, d)$ denote the number of cyclic subgroups of G of order d. Now define two equivalence relations on G as follows:

$$a \sim b \text{ if and only if } 0(a) = 0(b)$$
$$a \approx b \text{ if and only if } \langle a\rangle = \langle b\rangle$$

Then $a \approx b$ implies $a \sim b$ so each $\sim$ class is a union of $\approx$ classes, and there is (potentially) one $\sim$ class for each divisor d of n. Since each cyclic group of order d has $\varphi(d)$ generators, where φ is the Euler function, it follows that the $\sim$ class corresponding to d contains $\#(G, d)\varphi(d)$ elements. Hence we have

$$n = \sum_{d|n} \#(G, d)\varphi(d) \tag{1}$$

for any finite group G of order n. In particular, taking G to be the cyclic group of order n we have Gauss's formula

$$n = \sum_{d|n} \varphi(d) \tag{2}$$

Perhaps more importantly, by being exposed to our method, the students gain a much deeper understanding and appreciation of the uses of Gauss elimination. Our method unifies the ideas of two apparently unconnected processes and demonstrates the power of viewing matrix transformations as factorizations.

REFERENCES

1. Ben Noble and James W. Daniel, Applied Linear Algebra, 3rd edition, Prentice-Hall, Englewood Cliffs, N.J., 1988.
2. John R. Rice, Experiments on Gram-Schmidt orthogonalization, *Math. Comp.*, 20 (1966) 325–328.
3. Gilbert Strang, Linear Algebra and Its Applications, 3rd edition, Harcourt Brace Jovanovich, San Diego, 1988.

On Finite Groups and Finite Fields

J. D. Reid
Department of Mathematics, Wesleyan University, Middletown, CT 06457

Underlying one of the proofs, perhaps the most elementary one, of the fact that the multiplicative group of nonzero elements in a finite field is cyclic is an idea that we think perhaps deserves to be made more explicit. This can be done in terms of certain functions defined by finite groups, functions that might have some interest in their own right (see, for example, the *Remarks* below). Furthermore this approach to cyclic groups suggests an analogous treatment of a more purely field theoretic problem, namely the derivation of a formula for the number of irreducible polynomials of degree n over a finite field of cardinality q. While all this is essentially just a reformulation of standard results (cf. [1], pp. 259–261, or [3], p. 4, for example), we hope nonetheless that some readers will find this point of view interesting.

Groups. Suppose then that G is any group of order n. For elements a of G we denote by $0(a)$ the order of a and by $\langle a\rangle$ the cyclic subgroup of G generated by a. Further, let $\#(G, d)$ denote the number of cyclic subgroups of G of order d. Now define two equivalence relations on G as follows:

$$a \sim b \text{ if and only if } 0(a) = 0(b)$$
$$a \approx b \text{ if and only if } \langle a\rangle = \langle b\rangle$$

Then $a \approx b$ implies $a \sim b$ so each $\sim$ class is a union of $\approx$ classes, and there is (potentially) one $\sim$ class for each divisor d of n. Since each cyclic group of order d has $\varphi(d)$ generators, where φ is the Euler function, it follows that the $\sim$ class corresponding to d contains $\#(G, d)\varphi(d)$ elements. Hence we have

$$n = \sum_{d|n} \#(G, d)\varphi(d) \tag{1}$$

for any finite group G of order n. In particular, taking G to be the cyclic group of order n we have Gauss's formula

$$n = \sum_{d|n} \varphi(d) \tag{2}$$

since $\#(G, d) = 1$ for all d in this case. Subtracting (1) from (2) gives

$$0 = \sum_{d|n} [1 - \#(G, d)]\varphi(d). \tag{3}$$

The relations (1) and (3) are what we had in mind. As immediate consequences we have the two standard results:

THEOREM. *G is cyclic if and only if $\#(G, d) \leq 1$ for all d.*

Proof. If G is cyclic then $\#(G, d) = 1$ for all d. Conversely if $\#(G, d) \leq 1$ for all d then all coefficients of the $\varphi(d)$ in (3) are non-negative. Since the sum is zero, all coefficients must be zero. In particular, $\#(G, n) = 1$, i.e., G is cyclic. □

COROLLARY. *Any finite subgroup of the multiplicative group of a field is cyclic.*

Proof. If G is such a subgroup the existence of two distinct cyclic subgroups of G of order d would lead to more than d solutions to the equation $X^d - 1 = 0$ in the field, which is impossible, i.e., $\#(G, d) \leq 1$ for all d.

Remarks. It may be interesting to note that, for an arbitrary group G of order n, we obtain a parity of sorts between divisors d of n for which there is more than one cyclic subgroup of G of order d and those for which there are none. This is expressed in the following equation which comes directly from (3):

$$\sum_{\#(G,d)=0} \varphi(d) = \sum_{\#(G,d)>1} [\#(G, d) - 1]\varphi(d). \tag{4}$$

Secondly, from the structure theory of finite abelian groups it is clear that, for any abelian group, the function $\#(G, d)$ is a multiplicative function, i.e., that if d and d' are divisors of n and $gcd(d, d') = 1$, then

$$\#(G, dd') = \#(G, d)\#(G, d').$$

Fields. For a second application of this idea, let F be a finite field of cardinality q and let E be the extension of F of degree n. For a in E let $f_a(x)$ be the polynomial of least degree with coefficients in F having a as root and denote by $\deg f_a(x)$ the degree of this polynomial. Finally, let $N(q, d)$ denote the number of irreducible polynomials of degree d with coefficients in F. This depends only on q and d. Much as before define two equivalence relations on E by

$$a \sim b \text{ if and only if } \deg f_a(x) = \deg f_b(x).$$

$$a \approx b \text{ if and only if } f_a(x) = f_b(x).$$

Note that $a \sim b$ if and only if the fields $F(a)$ and $F(b)$ are equal. Again, $a \approx b$ implies $a \sim b$ and now there are $N(q, d)d$ elements in each $\sim$ class since each of the $N(q, d)$ irreducible polynomials of degree d have d (distinct) roots and distinct polynomials are relatively prime. Hence, we obtain an analog of (2), namely,

$$q^n = \sum_{d|n} N(q, d)d. \tag{5}$$

Now, in the usual way, Möbius inversion gives

$$N(q, n) = \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) q^d \tag{6}$$

where μ is the Möbius function. For another approach to the formula (5), and further results, see for example [**2**, p. 289].

Formulas (5) and (6), as well as (2), are due to Gauss. The author thanks the referee for informing him that the usual name for (2) is: Gauss's formula.

REFERENCES

1. I. N. Herstein, Abstract Algebra, Macmillan, New York, 1986.
2. Nathan Jacobson, Basic Algebra I, 2nd ed., W. H. Freeman, New York, 1985.
3. Jean-Pierre Serre, A Course in Arithmetic, Springer-Verlag, New York, 1973.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, MICHAEL A. FILASETA, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before November 30, 1991 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3445. *Proposed by Ronald A. Jansen and Pieter Moree, University of Leiden, The Netherlands.*

(a) Suppose p is an odd prime and $k \equiv 1 \pmod{p}$. Prove that for any positive integer n the highest power of p dividing n is equal to the highest power of p dividing $1 + k + k^2 + \cdots + k^{n-1}$.

(b) Suppose $k \equiv 1 \pmod{4}$. Prove that for any positive integer n the highest power of 2 dividing n is equal to the highest power of 2 dividing $1 + k + k^2 + \cdots + k^{n-1}$.

E 3446. *Proposed by Jean-Pierre Grivaux, Lycée Chaptal, Paris, France.*

Suppose $0 \leqslant u_0 < 1$ and $0 < \lambda < 1$. Define a sequence $\{u_n\}_{n=0}^{\infty}$ by putting

$$u_{n+1} = u_n + \lambda\sqrt{1 - u_n^2} \qquad (n = 0, 1, 2, \ldots).$$

For small n the sequence $\{u_n\}$ is real and increasing. Let $N = N(\lambda, u_0)$ be the first integer for which $u_N \geqslant 1$. For fixed u_0 find

$$\lim_{\lambda \to 0+} \lambda N(\lambda, u_0).$$

Solution by Kee-Wai Lau, Hong Kong. For $n \geq 2$ we show that $2 - L_n = n^{-1}\{\ln n + 1 - \ln 2 + O(\ln^2 n/n)\}$. The standard formula for arc length yields $L_n = \int_0^1 (1 + n^2 x^{2n-2})^{1/2}\,dx$. Letting $f(z) = 1 + z - (1 + z^2)^{1/2}$, this implies that $2 - L_n = \int_0^1 f(nx^{n-1})\,dx$. Setting $u = nx^{n-1}$, we can rewrite the preceding integral as $n^{1/(1-n)}(n-1)^{-1}(I + J)$, where

$$I = \int_0^n f(u)u^{-1}\,du \quad \text{and} \quad J = \int_0^n f(u)u^{-1}(u^{1/(n-1)} - 1)\,du.$$

Since $f(u)u^{-1}$ has the primitive $u - (1 + u^2)^{1/2} + \ln\{1 + (1 + u^2)^{1/2}\}$, we have

$$\begin{aligned} I &= n - (1 + n^2)^{1/2} + \ln\left\{1 + (1 + n^2)^{1/2}\right\} + 1 - \ln 2 \\ &= \ln n + 1 - \ln 2 + O(n^{-1}). \end{aligned}$$

Note also that $n^{1/(1-n)}(n-1)^{-1} = n^{-1} + O(\ln n/n^2)$.

To bound J, we break the integral at $u = 1$. Note that $f(u)u^{-1} < 1$ when $u < 1$. Since $\int_0^1 (1 - u^{1/(n-1)})\,du = n^{-1}$ and $(n^{1/(n-1)} - 1)\int_1^n f(u)u^{-1}\,du = O(\ln^2 n/n)$, we obtain the formula claimed.

Editorial comment. A few solvers obtained more precise results by more refined arguments. M. Ashbaugh showed that the terms described as $O(\ln^2 n/n)$ above are $-(\ln \frac{1}{2}n)^2/(2n) - (\pi^2 - 6)/(12n) + O(\ln^3 n/n^2)$. Results equivalent to $2 - L_n = \ln n/n + O(n^{-1})$ or stronger were obtained by those listed below; many different methods were used. Note that the inequality $2 - L_n < (\ln n)/n + 1/n$ is trivial, since it follows from the fact that L_n exceeds the length of the broken line with vertices $(0, 0)$, $(n^{-1/n}, n^{-1})$, and $(1, 1)$ and so is greater than $n^{-1/n} + 1 - n^{-1}$.

Solved also by J. Anglesio (France), M. Ashbaugh, D. Borwein (Canada), P. Bracken (Canada), E. R. Canfield, Z. Franco, M. Golomb, P. G. Kirmser, M. E. Kuczma (Poland), E. Levine, F. Levron (France), A. Meir (Canada), O. P. Lossers (Netherlands), S. Philipp, C. C. Rousseau & J. Warren, D. B. Tyler, the Chico Problem Group, an anonymous solver, and the proposer.

A Bijection between Matchings and Rooted Binary Trees

E 3349 [1989, 837]. *Proposed by László A. Székely, University of New Mexico, Albuquerque.*

Consider the set of matchings **M** on $2n$ labeled vertices and the set **T** of rooted binary trees with $n + 1$ labeled leaves (i.e., having n unlabeled nonleaf vertices and $2n$ edges). It is well known that $|\mathbf{M}| = (2n)!/(n!2^n)$. Prove that **T** has the same size by giving a constructive bijection between **M** and **T**.

Solution by Jerrold W. Grossman, Oakland University, Rochester, MI. The binary trees must be unordered in the sense that interchanging the subtrees of an internal vertex does not change the tree; this is the same as viewing the tree as a partially labeled graph. Assume that the leaves of the trees are labeled from 1 to $n + 1$ and the vertices to be matched are labeled from 1 to $2n$. We construct ϕ: $\mathbf{T} \to \mathbf{M}$ and ψ: $\mathbf{M} \to \mathbf{T}$. It will be clear from the construction that these are inverse functions.

Given $T \in \mathbf{T}$, we successively assign the labels $n + 2, \ldots, 2n$ to the nonleaf vertices other than the root as follows: having assigned labels up to $j - 1$, we assign j to the unlabeled vertex among those whose children are both labeled that has the child with the smallest label. Such a vertex exists, since we begin with all

leaves labeled and there are exactly $n - 1$ nonleaf vertices to be labeled. Define $\phi(T)$ to be the matching that pairs each vertex with its sibling.

Conversely, given a matching M, we successively assign to the vertices $n + 2$ through $2n$ a matched pair of children as follows: having assigned pairs to vertices up to $j - 1$, we assign to j the unassigned pair among those whose labels are both less than j that has the vertex with the smallest label. Such a pair exists, since when $j \leq 2n$ at most $2n - (j - 1)$ pairs are ineligible by having a number greater than $j - 1$ and exactly $(j - 1) - (n + 1)$ pairs are ineligible by being already assigned, leaving at least 1 of the n pairs available. Define $\psi(m)$ to be the tree such that the root has the only unassigned pair as its children and any other nonleaf vertex with label j has the pair assigned to j as its children.

Editorial comment. A more general version of this result will appear in a paper by the proposer.

Solved also by F. Bernhart, D. Callan, D. S. Franxblau, I. Gessel, R. J. Hendel, M. D. Hendy (New Zealand), K. M. Levasseur, D. Magagnosc, R. Stong, and the proposer.

A Sequence of Least Common Multiples

E 3350 [1989, 837]. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

For positive integers n and k with $1 \leq k \leq n$ let $A(n, k)$ denote the least common multiple of $n, n - 1, \ldots, n - k + 1$.

(a) Prove that 14 is the largest value of n for which there exists a positive integer $k \leq n$ such that

$$A(n,1) < A(n,2) < \cdots < A(n,k) = A(n,n),$$

i.e., 14 is the largest value of n for which $A(n, j)$ is strictly increasing in j until it reaches its maximum value $A(n, n)$.

(b) Let $f(n)$ be the largest value of k such that

$$A(n,1) < A(n,2) < \cdots < A(n,k).$$

Prove that $f(n) < C\sqrt{n}$ for some constant C.

(c) Prove that $f(n)$ tends to infinity with n. More precisely, prove that if $n > k! + k$, then $f(n) > k$.

Composite Solution by Richard Stong, University of California at Los Angeles, and John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. Observe that

$$A(n, k + 1) = \operatorname{lcm}\{n - k, A(n,k)\} \quad \text{for } 1 \leq k < n. \qquad (*)$$

Thus, $A(n, k) \leq A(n, k + 1)$ and equality holds if and only if $n - k$ divides $A(n, k)$. We first solve (b) by proving that $f(n) < 3\sqrt{n}$. Clearly, $f(1) = 1$. Note that for $m > 1$,

$$A(m^2, 2) = m^2(m^2 - 1) = (m^2 - m)(m^2 + m).$$

Thus, $m^2 - m$ divides $A(m^2, 2)$ and hence it also divides $A(m^2, m)$. So we have from (*) that $A(m^2, m) = A(m^2, m + 1)$. Hence, $f(m^2) \leq m$ for all positive integers m. Note that (*) implies that if $A(n, k) = A(n, k + 1)$, then $A(n + j, k + j)$

$= A(n + j, k + j + 1)$ for all positive integers j. Thus, $f(n + j) \leq f(n) + j$ for all positive integers n and j. Now, if $m^2 \leq n < (m + 1)^2$, then $n = m^2 + j$, where $0 \leq j \leq 2m$. Hence, $f(n) \leq f(m^2) + j \leq 3m \leq 3\sqrt{n}$, where one of the last two inequalities must in fact be a strict inequality.

To get the required proof for (a), first observe that

$$A(14, 1) < A(14, 2) < \cdots < A(14, 7) = A(14, 14).$$

Now, let n be an integer as in (a) with $n \geq 14$, and consider any prime $p \leq n$. Let k be the largest integer for which $p^k \leq n$, and let $m(p)$ be the largest multiple of p^k which is $\leq n$. It is easy to see that $A(n, n - m(p)) < A(n, n - m(p) + 1)$. Thus, for each prime $p \leq n$, we find that $m(p)$ must be in the set $\{n, n - 1, \ldots, n - f(n) + 1\}$. We now show that if p and q are distinct primes $\leq n$, then $m(p) \neq m(q)$. Assume otherwise. We may suppose that $q > p$. Then for some integers r, s, k, and t, we get that $m(p) = p^k r = q^t s = m(q)$ so that q divides r. But then $r \geq q > p$ implies a contradiction to the maximality of k. Hence, for the $\pi(n)$ primes $p \leq n$, we get distinct integers $m(p)$ in the set $\{n, n - 1, \ldots, n - f(n) + 1\}$. This implies that $f(n) \geq \pi(n)$. Recall that, from the result of part (b) already established, we have $f(n) < 3\sqrt{n}$. Thus, we get that $\pi(n) < 3\sqrt{n}$. The estimate $\pi(x) > x/\log x$ for all $x \geq 17$ (see (3.5) of J. B. Rosser and L. Schoenfeld, "Approximate formulas for some functions of prime numbers," *Illinois J. Math.* 6 (1962), 64–94) implies that $\pi(x) > 3\sqrt{x}$ for all $x \geq 289$. A direct computation shows that $\pi(x) > 3\sqrt{x}$ for all $x \in [179, 289)$. Thus, for (a), we need only consider $n \in (14, 179)$. These can readily be handled by using (*). For example, if $15 \leq n \leq 17$, observe that 9 does not divide $A(n, n - 9)$ and 10 divides $A(n, n - 10)$ so that $A(n, n - 10) = A(n, n - 9) < A(n, n)$. Thus, each of 15, 16, and 17 cannot be an integer n as in (a). Similarly, we can handle the cases $n \in [18, 179)$ by observing that for $18 \leq n \leq 21, 11 \nmid A(n, n - 11)$ and $12 | A(n, n - 12)$; for $22 \leq n \leq 25$, $13 \nmid A(n, n - 13)$ and $15 | A(n, n - 15)$; for $26 \leq n \leq 37, 19 \nmid A(n, n - 19)$ and $20 | A(n, n - 20)$; for $38 \leq n \leq 57, 29 \nmid A(n, n - 29)$ and $30 | A(n, n - 30)$; for $58 \leq n \leq 85, 43 \nmid A(n, n - 43)$ and $45 | A(n, n - 45)$; for $86 \leq n \leq 141, 71 \nmid A(n, n - 71)$ and $72 | A(n, n - 72)$; and for $142 \leq n \leq 178, 101 \nmid A(n, n - 101)$ and $102 | A(n, n - 102)$.

For (c), we use (*) to obtain

$$A(n, k + 1) = \operatorname{lcm}\{n - k, A(n, k)\} \geq \frac{A(n, k)(n - k)}{\prod_{j=1}^{k} \gcd(n - k, n - j + 1)}$$

$$\geq \frac{A(n, k)(n - k)}{k!}.$$

Therefore, if $n > k! + k$, then $A(n, 1) < \cdots < A(n, k) < A(n, k + 1)$ so that $f(n) > k$.

Solved also by the proposer. James Duemmel supplied solutions to parts (b) and (c).

Weighing All the Possibilities

E 3365 [1990, 64]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

Let S denote the sum of the integer weights that come with a two-pan balance scale. Give a necessary and sufficient condition on the weights that assures that all

integer-weighted objects in the range of 1 to S can be weighed exactly. Use of weights on both pans is permitted.

Solution by R. J. Chapman, University of Exeter, UK. Let $w_1 \le \cdots \le w_n$ be the weights, and let $S_j = \sum_{i=1}^{j} w_i$. Then every integer amount from 1 to S can be weighed if and only if $w_1 = 1$ and $w_{j+1} \le 2S_j + 1$ for $1 \le j < n$.

For sufficiency, we induct on j to show that every amount between 1 and S_j inclusive can be weighed using $\{w_1, \ldots, w_j\}$. This is plainly true for $j = 1$. If we can weigh i using $\{w_1, \ldots, w_j\}$, then we can also weigh $w_{j+1} - i$ and $w_{j+1} + i$ when w_{j+1} is also available, so we can weigh every amount from $w_{j+1} - S_j$ to $w_{j+1} + S_j = S_{j+1}$ using $\{w_1, \ldots, w_{j+1}\}$. Since $w_{j+1} - S_j \le S_j + 1$ by hypothesis, we can weigh every amount up to S_{j+1}.

For necessity, suppose we can weigh all amounts from 1 to S. The second largest weighable amount is $S - w_1$, which must be $S - 1$, so $w_1 = 1$. If $w_{j+1} > 2S_j + 1$ for some j, let $W = S - 2S_j - 1$; we claim W cannot be weighed. The largest amount that can be weighed without putting all of $\{w_{j+1}, \ldots, w_n\}$ in one pan is

$$\sum_{i=j+2}^{n} w_i + \sum_{i=1}^{j} w_i = S - w_{j+1} < W,$$

but the smallest amount that can be weighed using all of $\{w_{j+1}, \ldots, w_n\}$ in one pan is

$$\sum_{i=j+1}^{n} w_i - \sum_{i=1}^{j} w_i = S - 2S_j > W.$$

Solved also by D. M. Barrington, D. Callan, M. Falkowitz (Israel), R. Galia, M. Golomb, R. High, B. G. Klein, K. M. Levasseur, O. P. Lossers (The Netherlands), H. Noland, I. Schiermeyer (Germany), J. H. Steelman, R. Stong, D. P. Sumner, J. S. Sumner & K. L. Dove, H. H. Tehrani (Iran), A. Tissier (France), J. T. Ward, E. A.Weinstein, NSA Problems Group, University of Wyoming Problem Circle, and the proposer. One incorrect solution was received.

Heuristic vs. Rigorous

E 3380 [1990, 342]. *Proposed by Nick MacKinnon, Winchester College, U.K.*

Let N be the largest number of consecutive zeros occurring among the first 10^{1000} decimal places of $(1 + \sqrt{5})/2$.

(i) Give a heuristic argument for the order of magnitude of N to be expected.

(ii) Give a rigorous upper bound for N.

Solution by James G. Merickel (student), Temple University, Philadelphia, PA. (i) Assuming that the digits of $(1 + \sqrt{5})/2$ do not exhibit irregular statistical behavior, we would expect N to be around 1000. For in a randomly selected sequence of decimal digits a given string of 1000 digits occurs on average once every 10^{1000} digits. Moreover if $h > 0$ and if we consider the 10^{1000+h} possible strings of $1000 + h$ decimal digits, then fewer than $1/10^h$ of these strings can occur among the first 10^{1000} digits of a sequence of decimal digits, and so it would be reasonable to wager $10^h - 1$ to 1 against the occurrence of a particular one, such as the string of $1000 + h$ zeros.

integer-weighted objects in the range of 1 to S can be weighed exactly. Use of weights on both pans is permitted.

Solution by R. J. Chapman, University of Exeter, UK. Let $w_1 \le \cdots \le w_n$ be the weights, and let $S_j = \sum_{i=1}^{j} w_i$. Then every integer amount from 1 to S can be weighed if and only if $w_1 = 1$ and $w_{j+1} \le 2S_j + 1$ for $1 \le j < n$.

For sufficiency, we induct on j to show that every amount between 1 and S_j inclusive can be weighed using $\{w_1, \ldots, w_j\}$. This is plainly true for $j = 1$. If we can weigh i using $\{w_1, \ldots, w_j\}$, then we can also weigh $w_{j+1} - i$ and $w_{j+1} + i$ when w_{j+1} is also available, so we can weigh every amount from $w_{j+1} - S_j$ to $w_{j+1} + S_j = S_{j+1}$ using $\{w_1, \ldots, w_{j+1}\}$. Since $w_{j+1} - S_j \le S_j + 1$ by hypothesis, we can weigh every amount up to S_{j+1}.

For necessity, suppose we can weigh all amounts from 1 to S. The second largest weighable amount is $S - w_1$, which must be $S - 1$, so $w_1 = 1$. If $w_{j+1} > 2S_j + 1$ for some j, let $W = S - 2S_j - 1$; we claim W cannot be weighed. The largest amount that can be weighed without putting all of $\{w_{j+1}, \ldots, w_n\}$ in one pan is

$$\sum_{i=j+2}^{n} w_i + \sum_{i=1}^{j} w_i = S - w_{j+1} < W,$$

but the smallest amount that can be weighed using all of $\{w_{j+1}, \ldots, w_n\}$ in one pan is

$$\sum_{i=j+1}^{n} w_i - \sum_{i=1}^{j} w_i = S - 2S_j > W.$$

Solved also by D. M. Barrington, D. Callan, M. Falkowitz (Israel), R. Galia, M. Golomb, R. High, B. G. Klein, K. M. Levasseur, O. P. Lossers (The Netherlands), H. Noland, I. Schiermeyer (Germany), J. H. Steelman, R. Stong, D. P. Sumner, J. S. Sumner & K. L. Dove, H. H. Tehrani (Iran), A. Tissier (France), J. T. Ward, E. A.Weinstein, NSA Problems Group, University of Wyoming Problem Circle, and the proposer. One incorrect solution was received.

Heuristic vs. Rigorous

E 3380 [1990, 342]. *Proposed by Nick MacKinnon, Winchester College, U.K.*

Let N be the largest number of consecutive zeros occurring among the first 10^{1000} decimal places of $(1 + \sqrt{5})/2$.

(i) Give a heuristic argument for the order of magnitude of N to be expected.

(ii) Give a rigorous upper bound for N.

Solution by James G. Merickel (student), Temple University, Philadelphia, PA. (i) Assuming that the digits of $(1 + \sqrt{5})/2$ do not exhibit irregular statistical behavior, we would expect N to be around 1000. For in a randomly selected sequence of decimal digits a given string of 1000 digits occurs on average once every 10^{1000} digits. Moreover if $h > 0$ and if we consider the 10^{1000+h} possible strings of $1000 + h$ decimal digits, then fewer than $1/10^h$ of these strings can occur among the first 10^{1000} digits of a sequence of decimal digits, and so it would be reasonable to wager $10^h - 1$ to 1 against the occurrence of a particular one, such as the string of $1000 + h$ zeros.

(ii) We prove rigorously that $N \leq 5 \cdot 10^{999}$. Since

$$(1 + \sqrt{5})/2 = 1 + 1/\{(1 + \sqrt{5})/2\},$$

all partial quotients in the continued fraction expansion of $(1 + \sqrt{5})/2$ are 1. Thus it follows from the proof of Theorem 188 in [3] that for any positive integers p and q we have

$$|p/q - (1 + \sqrt{5})/2| > 1/(10q^2).$$

(In fact it is known that 10 can be replaced by $\sqrt{5} + \varepsilon$ for sufficiently large q. Cf. Theorem V of Chapter I of [1].) In particular

$$|p/10^n - (1 + \sqrt{5})/2| > 1/10^{2n+1}$$

for any positive integer n. This means that if the nth digit in the decimal expansion of $(1 + \sqrt{5})/2$ is non-zero, then no more than the next n consecutive digits can be zero. Thus a string of more than $5 \cdot 10^{999}$ consecutive zeros among the first 10^{1000} decimal digits of $(1 + \sqrt{5})/2$ is impossible.

Editorial comment. If N_i denotes the length of the longest string of consecutive zeros among the first 10^i decimal places of $(1 + \sqrt{5})/2$, it is easy to determine that $N_1 = 1$, $N_2 = 1$, and $N_3 = 3$. These data fit in with the conjecture made in the above solution that N_i has the same order of magnitude as i. Unfortunately it is not easy to extend these calculations as far as $N_{1000} = N$.

If r is around 1000, there is an approximate formula for the probability of at least one occurrence of r consecutive zeros among the first 10^{1000} places of a random sequence of decimal digits, that is, the probability of at least one success run of length r in 10^{1000} Bernoulli trials each with success probability $1/10$. In fact, according to the last problem of Chapter XIII of [2],

$$p_r \doteq 1 - \exp\{-10^{1000}(9/10)(1/10)^r\} = 1 - \exp\{-9 \cdot 10^{999-r}\}.$$

Thus $p_{999} \doteq 0.99988$, $p_{1000} \doteq 0.59343$. If h is a small positive integer, the probability p_{1000+h} of at least one occurrence of $1000 + h$ consecutive zeros among the first 10^{1000} places of a random sequence of decimal digits is around

$$1 - \exp\{-9 \cdot 10^{-h-1}\} \doteq 9 \cdot 10^{-h-1};$$

this agrees with the remark in the above solution of (*i*) that $p_{1000+h} < 10^{-h}$.

REFERENCES

1. J. W. S. Cassels, An Introduction to Diophantine Approximation, Cambridge Tract No. 45, 1957.
2. William Feller, An Introduction to Probability Theory and Its Applications, Volume I, Wiley, 1968.
3. G. H. Hardy and E. M.Wright, An Introduction to the Theory of Numbers, Oxford, 1960.

Solved also by the proposer.

SOLUTIONS OF ADVANCED PROBLEMS

Two Sets of Density One

6583 [1988, 880]. *Proposed by Moshe Laub, Jerusalem, Israel.*

(a) Let A be the set of positive integers z such that z^2 is expressible as a sum of two nonzero squares. Prove that A has asymptotic density one.

(b) Let B be the set of positive integers z such that the triangular number $T_z = z(z + 1)/2$ is expressible as a sum of two nonzero triangular numbers. Prove that B has asymptotic density one.

Solution by L. E. Mattics, University of South Alabama, Mobile. (a) We begin by showing that $z \in A$ if and only if z is divisible by at least one prime congruent to 1 modulo 4. On the one hand, if $z = jp$, where p is a prime congruent to 1 modulo 4, then $p = a^2 + b^2$, where $a > b > 0$, so that $p^2 = (a^2 - b^2)^2 + (2ab)^2$ and hence $z^2 = j^2p^2 = (a^2j - b^2j)^2 + (2abj)^2$. On the other hand, if z is not divisible by any prime congruent to 1 modulo 4, then there are exactly four solutions of $x^2 + y^2 = z^2$ in integers x, y, namely $x = \pm z$, $y = 0$ and $x = 0$, $y = \pm z$. (Cf. §5.10 of [3]).

Let C be the set of positive integers complementary to A, that is, C is the set of positive integers not divisible by any prime congruent to 1 modulo 4. Since the sum of the reciprocals of the primes congruent to 1 modulo 4 diverges, we see that C has natural density zero and so A has natural density one. (Cf. §§11.1 and 11.3 of [3].)

(b) We begin by showing that $z \in B$ if and only if $2z^2 + 2z + 1$ is composite. For $8T_z + 2 = (2z + 1)^2 + 1 = 2r$, where all prime factors of r are congruent to 1 modulo 4 (since -1 is a quadratic residue modulo any prime factor of r). Now $T_z = T_u + T_v$ for positive integers u and v if and only if

$$2r = (2u + 1)^2 + (2v + 1)^2. \qquad (*)$$

Since r has only prime factors congruent to 1 modulo 4, the number of solutions of $x^2 + y^2 = 2r$ in integers x, y (necessarily odd) is equal to $4d(r)$, where $d(r)$ is the number of positive integral divisors of r. The equality $2r = (2z + 1)^2 + 1$ provides eight solutions $x = \pm(2z + 1)$, $y = \pm 1$ and $x = \pm 1$, $y = \pm(2z + 1)$ of $x^2 + y^2 = 2r$ and so there is a solution of $(*)$ in positive integers u and v if and only if $d(r) > 2$, i.e., if and only if r is composite.

Let D be the set of positive integers complementary to B, that is, D is the set of positive integers z such that $2z^2 + 2z + 1$ is prime. Now for any prime q congruent to 1 modulo 4, $q < z$, the primality of $2z^2 + 2z + 1$ requires that z must avoid the two residue classes s modulo q such that

$$2(2s^2 + 2s + 1) = (2s + 1)^2 + 1 \equiv 0 \pmod{q}.$$

(If one of these two residue classes is represented by $s(q)$, where $0 < s(q) < q$, the other is represented by $q - 1 - s(q)$.) Thus if q_i is the ith prime congruent to 1 modulo 4, the density of D does not exceed

$$\prod_{i=1}^{m} \left(1 - \frac{2}{q_i}\right)$$

for any fixed m. Since the sum of the reciprocals of the primes congruent to 1 modulo 4 diverges, we see that D has natural density zero and so B has natural density one.

Editorial comment. It follows from Theorem 2.2 of [**1**] that if $G(x)$ is the number of elements of the set C not exceeding x and $H(x)$ is the number of elements of the set D not exceeding x, then

$$G(x) = O\big(x(\log x)^{-1/2}\big), \qquad H(x) = O\big(x(\log x)^{-1}\big).$$

The result of §183 of [**2**] shows that actually $G(x) \sim \alpha x(\log x)^{-1/2}$ as $x \to \infty$, where α is a positive constant. Heuristically we would also expect an asymptotic formula $H(x) \sim \beta x(\log x)^{-1}$ as $x \to \infty$, where β is a positive constant.

Thomas Maxsein and Paul R. Smith remarked that the equivalence "$z \in B$ if and only if $2z^2 + 2z + 1$ is composite" used in the solution of (b) has occurred at least twice in the literature, namely in [**4**] and [**5**].

REFERENCES

1. H. Halberstam and H.-E. Richert, Sieve Methods, Academic Press, 1974.
2. Edmund Landau, Handbuch der Lehre von der Verteilung der Primzahlen, Chelsea Publishing Co., 1953.
3. Ivan Niven and Herbert S. Zuckerman, An Introduction to the Theory of Numbers, Wiley, 1980.
4. Daniel Shanks, An analytic criterion for the existence of infinitely many primes of the form $(n^2 + 1)/2$, *Illinois J. Math.*, 8 (1964) 377–379.
5. Waclaw Sierpiński, Sur les nombres triangulaire qui sont sommes de deux nombres triangulaires, *Elem. Math.*, 17 (1962) 63–65.

Solved also by the late Emil Grosswald, Thomas Maxsein & Paul R. Smith, and the editors. The Grosswald and Maxsein-Smith solutions were similar to that given above. The editors' solution used the method of Orrin Frink, Almost Pythogorean triples, *Math. Magazine*, 60 (1987) 234–236.

A Property of the Rationals

6614 [1989, 846]. *Proposed by Justin G. MacCarthy, Deming, NM.*

A proper subset A of an infinite abelian group G is called a "wedge" in G if the family of all translates $A + x$, $x \in G$, is closed under finite intersection. Prove that in Q^+, the additive group of the rationals, the only wedges are half-lines (open or closed).

Solution by Richard Stong, University of California at Los Angeles. Let A be a wedge in Q^+. Define $B = \{x \in Q^+: A + x \subseteq A\}$. We will show that $B = Q \cap [0, \infty)$ or $B = Q \cap (-\infty, 0]$. This clearly implies that A is a half-line.

First we will show B is a wedge. Clearly $0 \in B$, so B is nonempty and $B \neq Q^+$, since otherwise we would have $A = Q^+$. Suppose $y \in Q^+$. Then clearly $B + y = \{x \in Q^+: A + x \subseteq A + y\}$. Thus since A is a wedge

$$\bigcap_{i=1}^{n} (B + y_i) = \left\{x \in Q^+: A + x \subseteq \bigcap_{i=1}^{n} (A + y_i)\right\}$$
$$= \{x \in Q^+: A + x \subseteq A + z\} = B + z,$$

for some z. That is, B is a wedge. Next note that B is closed under addition, i.e. B

is a submonoid of Q^+. This is clear since $A + x \subseteq A$ and $A + y \subseteq A$ clearly implies $A + x + y \subseteq A + y \subseteq A$.

Suppose B contained both positive and negative elements, say a and $-b$. Since B is closed under addition B would contain their positive multiples and hence would contain c and $-c$ for some positive c. Rescaling A, and hence B, we may assume $\pm 1 \in B$. Since B is a wedge, for any integer n we have

$$B \cap \left(B + \frac{1}{n}\right) = B + \frac{p}{q},$$

for some $p/q \in Q^+$ (we may assume $p > 0$ by, if necessary, replacing A by $-A$). Hence p/q and $p/q - 1/n$ belong to B. Since B is closed under addition, we would thus have

$$-\frac{1}{n} = (q-1)\frac{p}{q} + \left(\frac{p}{q} - \frac{1}{n}\right) + p(-1) \in B$$

and

$$\frac{1}{n} = (nq - n + 1)\frac{p}{q} + (n-1)\left(\frac{p}{q} - \frac{1}{n}\right) + np(-1) + 1 \in B.$$

Since B is closed under addition, this would imply $B = Q^+$.

Therefore B contains elements of only one sign. Replacing A by $-A$, if necessary, we may assume $B \subseteq Q \cap [0, \infty)$. Fix $x \in Q \cap [0, \infty)$, let y be such that $B \cap (B + x) = B + y$, and let k be the least positive integer such that there is a nonnegative integer n with $ny + kx \in B$ (k exists since for some large K and M, $Kx = My \in B$). Then

$$B + (n+1)y + (k-1)x = (B + ny + kx) \cap (B + ny + (k-1)x) \subseteq B,$$

since $B + ny + kx \subseteq B$. In particular, $(n+1)y + (k-1)x \in B$. This contradicts the minimality of k unless $k = 1$. Thus there is a least nonnegative integer n for which $ny + x \in B$. However, since $B \cap (B + x) = B + y$ and $ny \in B$, this would imply $ny + x \in B + y$ or $(n-1)y + x \in B$. This contradicts the minimality of n unless $n = 0$ and hence $x \in B$. Therefore $B = Q \cap [0, \infty)$, as desired.

Editorial comment. Duane Broline and the proposer point out that the result is false for the additive reals (just subject the half-line wedge $0 \leqslant x$ to a totally discontinuous linear transformation—this yields an everywhere dense image wedge). On the other hand, F. J. Flanigan proved the result for an arbitrary subgroup of the additive rationals.

Solved also by Duane Broline, F. J. Flanigan, and the proposer.

Averages of the GCD and of the LCM

6615 [1989, 847]. *Proposed by Kee-Wai Lau, Hong Kong.*

Denote by (i, j) and $[i, j]$ respectively the greatest common divisor and the least common multiple of the positive integers i and j. Prove that as n tends to infinity

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (i, j) \sim \frac{n^2 \log n}{\zeta(2)}, \tag{i}$$

is a submonoid of Q^+. This is clear since $A + x \subseteq A$ and $A + y \subseteq A$ clearly implies $A + x + y \subseteq A + y \subseteq A$.

Suppose B contained both positive and negative elements, say a and $-b$. Since B is closed under addition B would contain their positive multiples and hence would contain c and $-c$ for some positive c. Rescaling A, and hence B, we may assume $\pm 1 \in B$. Since B is a wedge, for any integer n we have

$$B \cap \left(B + \frac{1}{n}\right) = B + \frac{p}{q},$$

for some $p/q \in Q^+$ (we may assume $p > 0$ by, if necessary, replacing A by $-A$). Hence p/q and $p/q - 1/n$ belong to B. Since B is closed under addition, we would thus have

$$-\frac{1}{n} = (q-1)\frac{p}{q} + \left(\frac{p}{q} - \frac{1}{n}\right) + p(-1) \in B$$

and

$$\frac{1}{n} = (nq - n + 1)\frac{p}{q} + (n-1)\left(\frac{p}{q} - \frac{1}{n}\right) + np(-1) + 1 \in B.$$

Since B is closed under addition, this would imply $B = Q^+$.

Therefore B contains elements of only one sign. Replacing A by $-A$, if necessary, we may assume $B \subseteq Q \cap [0, \infty)$. Fix $x \in Q \cap [0, \infty)$, let y be such that $B \cap (B + x) = B + y$, and let k be the least positive integer such that there is a nonnegative integer n with $ny + kx \in B$ (k exists since for some large K and M, $Kx = My \in B$). Then

$$B + (n+1)y + (k-1)x = (B + ny + kx) \cap (B + ny + (k-1)x) \subseteq B,$$

since $B + ny + kx \subseteq B$. In particular, $(n+1)y + (k-1)x \in B$. This contradicts the minimality of k unless $k = 1$. Thus there is a least nonnegative integer n for which $ny + x \in B$. However, since $B \cap (B + x) = B + y$ and $ny \in B$, this would imply $ny + x \in B + y$ or $(n-1)y + x \in B$. This contradicts the minimality of n unless $n = 0$ and hence $x \in B$. Therefore $B = Q \cap [0, \infty)$, as desired.

Editorial comment. Duane Broline and the proposer point out that the result is false for the additive reals (just subject the half-line wedge $0 \leqslant x$ to a totally discontinuous linear transformation—this yields an everywhere dense image wedge). On the other hand, F. J. Flanigan proved the result for an arbitrary subgroup of the additive rationals.

Solved also by Duane Broline, F. J. Flanigan, and the proposer.

Averages of the GCD and of the LCM

6615 [1989, 847]. *Proposed by Kee-Wai Lau, Hong Kong.*

Denote by (i, j) and $[i, j]$ respectively the greatest common divisor and the least common multiple of the positive integers i and j. Prove that as n tends to infinity

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(i,j) \sim \frac{n^2 \log n}{\zeta(2)}, \tag{i}$$

Using (2), (3) and (4) we obtain

$$\sum_{j=1}^{n}\sum_{i=1}^{n}(i,j) = \frac{n^2}{\zeta(2)}\sum_{j\leqslant n}\frac{1}{j} + O\left(n\sum_{j\leqslant n}\log\frac{n}{j}\right) + O(n^2)$$

$$= \frac{n^2}{\zeta(2)}\log n + O(n^2),$$

which proves (A).

The proof of (B) uses the arithmetical function

$$h(m) = \sum_{d|m} d\mu(d),$$

where μ is the Möbius function. Note that $h(m) = \prod_{p|m}(1-p)$, so that

$$|h(m)| = \prod_{p|m}(p-1) < \prod_{p|m} p \leqslant m.$$

Möbius inversion gives us

$$\frac{1}{m} = \sum_{d|m}\frac{h(d)}{d}.$$

Take $m = (i,j)$ in this formula, note that $[i,j] = ij/(i,j)$, and sum on i to obtain

$$\sum_{i=1}^{n}[i,j] = \sum_{i=1}^{n} ij \sum_{d|(i,j)}\frac{h(d)}{d} = j\sum_{d|j}\sum_{q=1}^{\lfloor n/d\rfloor}(qd)\frac{h(d)}{d}$$

$$= \frac{j}{2}\sum_{d|j}h(d)\left\{\left(\frac{n}{d}\right)^2 + O\left(\frac{n}{d}\right)\right\} = \frac{jn^2}{2}\sum_{d|j}\frac{h(d)}{d^2} + O\left(\frac{jn}{2}\sum_{d|j}\frac{h(d)}{d}\right)$$

$$= \frac{jn^2}{2}\sum_{d|j}\frac{h(d)}{d^2} + O(nj\tau(j)),$$

where $\tau(j)$ is the number of positive integers dividing j. Summing on j and using Dirichlet's estimate for $\sum\tau(j)$, we find

$$\sum_{j=1}^{n}\sum_{i=1}^{n}[i,j] = \frac{n^2}{2}\sum_{j=1}^{n} j\sum_{d|j}\frac{h(d)}{d^2} + O(n^3\log n)$$

$$= \frac{n^2}{2}\sum_{qd\leqslant n}(qd)\frac{h(d)}{d^2} + O(n^3\log n)$$

$$= \frac{n^2}{2}\sum_{d=1}^{n}\frac{h(d)}{d}\sum_{q\leqslant n/d} q + O(n^3\log n)$$

$$= \frac{n^4}{4}\sum_{m=1}^{n}\frac{h(m)}{m^3} + O\left(n^3\sum_{m=1}^{n}\frac{|h(m)|}{m^2}\right) + O(n^3\log n)$$

$$= \frac{n^4}{4}\sum_{m=1}^{n}\frac{h(m)}{m^3} + O\left(n^3\sum_{m=1}^{n}\frac{1}{m}\right) + O(n^3\log n)$$

$$= \frac{n^4}{4}\sum_{m=1}^{n}\frac{h(m)}{m^3} + O(n^3\log n).$$

The principal term is

$$\frac{n^4}{4}\sum_{m=1}^{n}\frac{h(m)}{m^3} = \frac{n^4}{4}\sum_{m=1}^{n}\frac{1}{m^3}\sum_{d\mid m} d\mu(d) = \frac{n^4}{4}\sum_{qd\le n}\frac{1}{(qd)^3}\,d\mu(d)$$

$$= \frac{n^4}{4}\sum_{d=1}^{n}\frac{\mu(d)}{d^2}\left\{\zeta(3) + O\left(\frac{d^2}{n^2}\right)\right\}$$

$$= \frac{n^4}{4}\zeta(3)\sum_{d=1}^{n}\frac{\mu(d)}{d^2} + O(n^3)$$

$$= \frac{n^4\zeta(3)}{4\zeta(2)} + O(n^3),$$

from which we obtain (B).

Editorial comment. A number of solvers found various error terms and generalizations. In particular, a collection of results on these sums and related ones may be found in Persi Diaconis and Paul Erdös, "On the distribution of the greatest common divisor," Technical Report No. 12, Department of Statistics, Stanford University, Stanford, 1977. They trace results of this type back to E. Cesàro, Etude moyenne du plus grand commun diviseur de deux nombres," *Ann. Mat. Pura Appl.*, (2) 13 (1885) 235–250.

Solved also by Persi Diaconis and Paul Erdös (jointly), Mark C. W. Jones (U.K.), Haseo Ki (Korea), Enoch Kin-S. Lee, O. P. Lossers (The Netherlands), Richard Stong, Alain Tissier (France), T. P. T. Williams, and the proposer.

A Formula for the mth Largest Prime Factor of an Integer

6618 [1989, 942]. *Proposed by Armel Mercier, Université du Québec à Chicoutimi, Canada.*

Let $\omega(n)$ denote the number of distinct prime factors of n, $p(n)$ the smallest prime factor of n, and $P_m(n)$ the mth largest prime factor of n.

(i) Prove that for any nonnegative integer m we have

$$P_{m+1}(n) = (-1)^{m+1}\sum_{d\mid n}\binom{\omega(d)-1}{m}\mu(d)p(d),$$

where μ denotes the Möbius function.

(ii) Prove that for any nonnegative integer m we have

$$\sum_{d\mid n}\{\omega(d)\}^m\mu(d)p(d) = \sum_{j=0}^{m}(-1)^{j+1}j!S(m+1,j+1)P_{j+1}(n),$$

where $S(m, j)$ denotes the Stirling number of the second kind.

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. It should probably be noted that these formulas assume that $P_m(n) = 0$ if $m > \omega(n)$ and that $p(1) = 0$.

(i) We establish the desired formula by induction on $\omega(n)$. If $\omega(n) = 0$ (i.e., $n = 1$), then trivially

$$(-1)^{m+1} \sum_{d|n} \binom{\omega(d) - 1}{m} \mu(d)p(d) = 0 = P_{m+1}(n)$$

for all nonnegative integers m.

So assume that the formula has been established for all positive integers having fewer than $\omega(n)$ distinct prime factors. We can factor n as $n = aq^r$ where q is a prime larger than any prime factor of a. For the case $m = 0$, note that

$$\begin{aligned}
&(-1) \sum_{d|n} \binom{\omega(d) - 1}{0} \mu(d)p(d) \\
&\quad = (-1) \sum_{d|a} \sum_{j=0}^{r} \binom{\omega(dq^j) - 1}{0} \mu(dq^j)p(dq^j) \\
&\quad = (-1) \left[\sum_{j=1}^{r} \mu(q^j)q + \sum_{\substack{d|a \\ d>1}} \{\mu(d)p(d) - \mu(d)p(d)\} \right] \\
&\quad = q = P_1(n).
\end{aligned}$$

If $m > 0$ we have

$$\begin{aligned}
&(-1)^{m+1} \sum_{d|n} \binom{\omega(d) - 1}{m} \mu(d)p(d) \\
&\quad = (-1)^{m+1} \sum_{d|a} \sum_{j=0}^{r} \binom{\omega(dq^j) - 1}{m} \mu(dq^j)p(dq^j) \\
&\quad = (-1)^{m+1} \sum_{\substack{d|a \\ d>1}} \left[\binom{\omega(d) - 1}{m} - \binom{\omega(d)}{m} \right] \mu(d)p(d) \\
&\quad = (-1)^{m} \sum_{d|a} \binom{\omega(d) - 1}{m - 1} \mu(d)p(d) \\
&\quad = P_m(a) = P_{m+1}(n).
\end{aligned}$$

(ii) We require the following basic property of the Stirling numbers of the second kind (which can be used as their defining property): If m is a non-negative integer and x is an indeterminate, then

$$x^m = \sum_{j=0}^{m} S(m + 1, j + 1) \prod_{i=1}^{j} (x - i), \qquad (*)$$

where the empty product (when $j = 0$) is understood to be 1.

The result of (ii) can be deduced from (*) and the result of (i) as follows:

$$\sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1) P_{j+1}(n)$$

$$= \sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1)(-1)^{j+1} \sum_{d|n} \binom{\omega(d)-1}{j} \mu(d) p(d)$$

$$= \sum_{d|n} \sum_{j=0}^{m} j! S(m+1, j+1) \binom{\omega(d)-1}{j} \mu(d) p(d)$$

$$= \sum_{d|n} \{\omega(d)\}^{m} \mu(d) p(d).$$

Solved also by David Callan, Haseo Ki (Korea), and the proposer.

The Ratio of the Roots of an Irreducible Quadratic

6621 [1990, 76]. *Proposed by William C. Waterhouse, Pennsylvania State University, University Park.*

Which numbers can occur as a ratio of roots of an irreducible quadratic equation? More precisely, suppose K is an algebraically closed field. For each subfield E of K and each separable quadratic extension F of E with conjugation $\bar{}$, certain elements of K occur as ratios $\bar{c}/c$ with c in F. Which elements of K (besides 1) do arise in this way when we allow all possible choices of E and F?

Solution by Dennis R. Estes, University of Southern California, Los Angeles. The problem is to determine the elements θ (other than 0 and 1) in an algebraically closed field K having property

A. There exist subfields E, F of K with F separable of degree 2 over E, and $c \in F$ such that $\theta = c^*/c$ where * is the generator of the Galois group of F over E.

Let Π denote the subfield of K generated by 1. We shall show that A is equivalent to

B. Either θ is transcendental over Π, or θ and θ^{-1} are roots of the same irreducible polynomial over Π.

It is easy to see that B follows from A. Assume θ is algebraic over Π with minimal polynomial $p(x) \in \Pi[x]$; since * fixes $p(x)$ we have $p(\theta) = p(\theta^*) = p(\theta^{-1}) = 0$.

To show that A follows from B, first consider the case $\theta \neq -1$. Observe that B implies the existence of an automorphism * of $F = \Pi(\theta)$ sending θ to θ^{-1}. Let E denote the fixed field of *. Since $\theta \neq \theta^{-1} = \theta^*$, the generator * has order 2 and by elementary Galois theory F is a separable degree 2 extension of E with Galois group generated by *. (Alternatively, $\theta + \theta^{-1} \in E$ and since θ is a root of

$$g(x) = x^2 - (\theta + \theta^{-1})x + 1 = 0,$$

we have $[F : E] \leqslant 2$. Since $\theta \notin E$, we have $[E : F] = 2$ and since $g(x)$ has

The result of (ii) can be deduced from (*) and the result of (i) as follows:

$$\sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1) P_{j+1}(n)$$

$$= \sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1)(-1)^{j+1} \sum_{d|n} \binom{\omega(d)-1}{j} \mu(d) p(d)$$

$$= \sum_{d|n} \sum_{j=0}^{m} j! S(m+1, j+1) \binom{\omega(d)-1}{j} \mu(d) p(d)$$

$$= \sum_{d|n} \{\omega(d)\}^{m} \mu(d) p(d).$$

Solved also by David Callan, Haseo Ki (Korea), and the proposer.

The Ratio of the Roots of an Irreducible Quadratic

6621 [1990, 76]. *Proposed by William C. Waterhouse, Pennsylvania State University, University Park.*

Which numbers can occur as a ratio of roots of an irreducible quadratic equation? More precisely, suppose K is an algebraically closed field. For each subfield E of K and each separable quadratic extension F of E with conjugation $\bar{}$, certain elements of K occur as ratios $\bar{c}/c$ with c in F. Which elements of K (besides 1) do arise in this way when we allow all possible choices of E and F?

Solution by Dennis R. Estes, University of Southern California, Los Angeles. The problem is to determine the elements θ (other than 0 and 1) in an algebraically closed field K having property

A. There exist subfields E, F of K with F separable of degree 2 over E, and $c \in F$ such that $\theta = c^*/c$ where * is the generator of the Galois group of F over E.

Let Π denote the subfield of K generated by 1. We shall show that A is equivalent to

B. Either θ is transcendental over Π, or θ and θ^{-1} are roots of the same irreducible polynomial over Π.

It is easy to see that B follows from A. Assume θ is algebraic over Π with minimal polynomial $p(x) \in \Pi[x]$; since * fixes $p(x)$ we have $p(\theta) = p(\theta^*) = p(\theta^{-1}) = 0$.

To show that A follows from B, first consider the case $\theta \neq -1$. Observe that B implies the existence of an automorphism * of $F = \Pi(\theta)$ sending θ to θ^{-1}. Let E denote the fixed field of *. Since $\theta \neq \theta^{-1} = \theta^*$, the generator * has order 2 and by elementary Galois theory F is a separable degree 2 extension of E with Galois group generated by *. (Alternatively, $\theta + \theta^{-1} \in E$ and since θ is a root of

$$g(x) = x^2 - (\theta + \theta^{-1})x + 1 = 0,$$

we have $[F : E] \leqslant 2$. Since $\theta \notin E$, we have $[E : F] = 2$ and since $g(x)$ has

precisely θ and θ^{-1} as its distinct roots, F is separable over E.) Now observe that

$$\theta = \frac{(1 + \theta^{-1})^*}{1 + \theta^{-1}}.$$

(One could also handle this case with Hilbert's Theorem 90).

We now consider the case $\theta = -1$. Since $\theta \neq 1$ the field K does not have characteristic 2. Hence either Π is a field with p elements where p is an odd prime, or Π is isomorphic to the field of rationals. In either case $\Pi \neq \Pi^2$. Take $a \in \Pi$ with $a \notin \Pi^2$. Then

$$-1 = \frac{(\sqrt{a})^*}{\sqrt{a}},$$

where * is the generator of the Galois group of the separable extension $\Pi(\sqrt{a})/\Pi$.

Solved also by R. J. Chapman (U.K.) and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Mathematics and Its History. By John Stillwell. Springer-Verlag, 1989, 10 + 371 pp.

JEFFREY NUNEMACHER
Department of Mathematical Sciences, Ohio Wesleyan University, Delaware, OH 43015

As we wrestle with the revision of the mathematics major in response to the impetus of technological change and declining numbers of students, we face a variety of new curricular problems. But there are also some older problems which have never been successfully resolved. First, there are simply too many topics that need to be included in the training of any student who is serious about mathematics. This problem has become particularly acute with the recent much-needed addition to the curriculum of courses in discrete mathematics and modeling. Our programs are structured in such a way that a student either takes a full course on a subject or else learns nothing about it at all. As a result, students often remain totally ignorant of such venerable but still important subjects as number theory, calculus of variations, and geometry—projective, differential, or hyperbolic.

A second problem is that most students majoring in mathematics, even very good ones who take many courses in the discipline, complete their undergraduate programs with little sense of the sweep and unity of the subject. How can they develop such an overview when we offer them a hodgepodge of almost unrelated courses? It is rare to see any calculus or geometry used in an algebra course or to encounter a group or a matrix in undergraduate analysis. Even for those few students who continue on to graduate school and into the profession, this lack of perspective is rarely remedied. In too many graduate programs once the utilitarian basic training is completed, there is little encouragement to take or to audit courses away from a narrow research area.

It has been proposed that an undergraduate course in the history of mathematics, perhaps as a capstone course in the major, might partially solve these problems. But few of us have more than a rudimentary knowledge of the history of our subject or much of a desire to expend energy in learning it. One can also argue that a serious student is better off learning some new subject in depth than taking a course in mathematical history, which necessarily must be somewhat superficial. Thus such a course would be best for our less dedicated or weaker students who probably don't have the necessary background or maturity to appreciate it anyway! For these reasons I have never felt very enthusiastic about this or any other capstone idea. But Stillwell's book is making me think again. It is the first book concerned with the history of mathematics that I have found really captivating. This is partly because it uses a historical perspective to discuss a broad range of mathematical topics and in the process fills many of the gaps mentioned above.

Let me illustrate the advantages of the historical point of view with two geometric examples, one with an algebraic flavor and the other from differential

geometry. We all learn about lines and conics in calculus courses, but for most of us the study of algebraic curves stops with degree two. What about cubics, the next most complicated case? Newton was the first to attempt a classification of cubics similar to that for conics as ellipses, parabolas, and hyperbolas. He wrote several papers on the topic over a period of twenty-five years, the last in 1695. He wished to classify according to their asymptotic behavior all curves defined by the general nondegenerate cubic equation in two real variables

$$ay^3 + bxy^2 + cx^2y + dx^3 + ey^2 + fxy + gx^2 + hy + kx + l = 0.$$

Here nondegenerate means that the equation does not factor over R, so that its graph contains no components which are lines or conics. The rotations and translations that simplify conics do not suffice for the more complicated set of cubics. But by employing a general affine change of coordinates Newton was able to reduce a nondegenerate cubic to one of four types:

$$y = Ax^3 + Bx^2 + Cx + D \tag{1}$$

$$xy = Ax^3 + Bx^2 + Cx + D \tag{2}$$

$$y^2 = Ax^3 + Bx^2 + Cx + D \tag{3}$$

$$Mxy^2 + Ny = Ax^3 + Bx^2 + Cx + D, \qquad M \neq 0. \tag{4}$$

To do this, he performed a tour de force of classical algebra which involved passing through an equation with 84 terms. Today it is much less difficult to carry out this simplification. For a hint on how to do it efficiently (as well as for a lucid modern treatment of the theory of curves), see [**2**, p. 91]. Notice that by allowing a general affine change of coordinates, Newton enlarged the group of permissible transformations beyond the Euclidean group, which suffices to classify the conics.

Newton then proceeded to divide these curves into 72 species according to their asymptotic behaviour. Actually, there should be 78 species in all according to his classification scheme—Newton missed 6 species which were found by later authors. Types (1) and (2), which yield one species each, are curves that we all graph in elementary calculus. There are five species of type (3), depending on whether the right-hand side of the equation has three distinct real roots, two complex and one real one, a single real root of multiplicity three, or a real root of multiplicity two either greater or less than the other root of multiplicity one. It is a good exercise for calculus students to obtain this five-fold breakdown and to draw a graph of each species. Type (4), which yields the remaining 71 species, is not easily dealt with by hand (though of course Newton managed it!), but representative curves can be obtained with *Derive* or *Mathematica*, since one can complete the square in y and thus express y in terms of x. For those looking for an interesting exploratory exercise using such a package, here it is. Newton presented a graph of each species (see [**4**], [**5**], or [**2**, pp. 93–98], where a portion of Newton's article is reproduced).

Newton's classification scheme was criticized by Euler and others because his geometric criteria appeared to lack a unifying theme and resulted in so many different species. Later authors were undeterred however: Plücker in 1835 gave a more refined classification and obtained 219 different species. For a discussion of Newton's and Plücker's criteria see [**3**]. But the real answer to Euler's criticism was already hinted at in a brief section of Newton's paper entitled "On the Genesis of

Curves by Shadows". Newton asserted that all nondegenerate cubic curves have as a projective image one of the five species of type (3). Thus by adopting the point of view of projective geometry or, equivalently, by enlarging the group of permissible transformations from the affine group to the projective group $\mathbf{GL}(3,\mathbf{R})/\{k \cdot Id\}$, the classification of cubics simplifies greatly.

Considering the classification problem within the projective plane makes good intuitive sense, since that plane is formed by adding to the affine plane a point for each asymptotic direction, and it was the variety of asymptotic behaviour which gave rise to the large number of species of cubics. It was not until the eighteenth century that projective geometry was sufficiently well developed to provide a rigorous proof of Newton's claim (see [1], pp. 122–124). The final step in the classification of cubics, a contribution of nineteenth century mathematics, was to introduce the complex projective plane with its associated larger group of complex projective transformations $\mathbf{GL}(3,\mathbf{C})/\{k \cdot Id\}$. Since the field of complex numbers is algebraically closed, the factorization of polynomials simplifies and all cubics fall into just three classes according to the number of distinct complex roots of the right-hand side of (3). In the case of a single root the cubic is projectively equivalent to Neil's parabola $y^2 = x^3$, which has a cusp at the origin, and in the case of exactly two roots, one having multiplicity two, to the curve $y^2 = x^2(x + 1)$, which has a node. Thus in both of these cases the curve has a singularity. In the remaining case of three distinct roots the curve is everywhere nonsingular and is equivalent to an equation of the form $y^2 = 4x^3 - g_2x - g_3$ with $g_2^3 - 27g_3^2 \neq 0$, which is known as the Weierstrass normal form. Two such curves are projectively equivalent if and only if their J-invariants $J = g_2^3/(g_2^3 - 27g_3^2)$ are the same. For a full discussion of this classification see [2, pp. 283–305].

There are essential topological and algebraic differences between the singular and nonsingular cases. Singular cubic curves when viewed as surfaces in two-dimensional complex projective space have genus zero and possess a parametrization by rational functions, while nonsingular curves have genus one and admit no rational parametrization. So how can a nonsingular curve be parametrized? By elliptic functions, and this leads us to yet another topic of nineteenth century mathematics. The genus one curves are called elliptic curves and are currently the object of intensive study because of their connections with such arithmetic phenomena as Fermat's Last Theorem. By tracing the history of cubic curves we have seen the application of tools from calculus, linear algebra, group theory, topology, projective geometry, and complex variables, all to one very concrete problem. What better way is there to see that all of mathematics fits together?

Our second example concerns the origins of the Gauss-Bonnet theorem, a fundamental result in differential geometry which links the curvature of a surface to its topological type. Although this theorem can be appreciated with only a smattering of differential geometry and topology, it is usually encountered only at the end of a full course in these subjects. One version of this theorem asserts that if S is a two-dimensional surface in $\mathbf{R}^3$ which is bounded by a finite number of smooth curves C_i, then

$$\sum \alpha_j + \sum \int_{C_i} \kappa_g \, ds + \int_S \int K \, dA = 2\pi\chi(S), \tag{5}$$

In this formula K denotes the Gaussian curvature function on the surface which at

each point measures the bending of S, and κ_g denotes the geodesic curvature function on the boundary curves C_i which is the projection of the usual curvature κ of a three-dimensional curve onto the tangent plane of the surface. Thus, in a planar region κ_g agrees with curvature as defined in calculus. The α_j (if present) are the (signed) exterior angles at which the curves C_i meet, and $\chi(S)$ denotes the Euler characteristic of the surface S, which can be calculated from the familiar formula $\chi(S) = V - E + F$. For precise definitions of these concepts see any book on differential geometry (or the book being reviewed here). The important fact is that the curvature functions capture the geometry of the surface S and the curves C_i, while the Euler characteristic determines its topology. In the form stated here the theorem is a late nineteenth century elaboration of work of Gauss in 1825 and Bonnet in 1848, but it had some very early antecedents which help considerably in appreciating the theorem and in making it plausible.

Perhaps the simplest and earliest recognizable precursor is the familiar theorem in Euclidean geometry that the sum of the exterior angles of a convex polygon is equal to 2π. Since a polygon is bounded by straight lines and lies in a plane, both curvature terms are zero. Its Euler characteristic can be directly calculated to be one, and we see that this ancient result is a special case of the Gauss-Bonnet theorem. Notice that the angular terms in (5) are a reasonable measure of turning at those points where the geodesic curvature is not defined. Thus the Euclidean theorem can be interpreted as asserting that the total turning in traversing the boundary of a convex polygon is 2π.

Another simple special case is the observation that for a simple closed curve in the plane $\int \kappa \, ds = 2\pi$. If curvature κ is calculated by the formula $d\theta/ds$, where at each point on the curve θ is the polar angle and s denotes arc length, this result follows directly from the Fundamental Theorem of Calculus.

A third antecedent is due to Thomas Harriot, a colorful figure in English mathematics who, undeservedly, is little known today. He was a protegé of Sir Walter Raleigh and wrote a book in 1588 about Raleigh's first attempt to establish a settlement in Virginia. Harriot discovered several particular facts which have had a significant impact on the development of mathematics and which should have earned him a permanent place in mathematical history. One of these, which is sketched in Stillwell's book, is an elegant geometric relation between the area, radius, and angles of a spherical triangle. A spherical triangle is a triangle on the surface of a sphere whose sides are formed by arcs of great circles. Harriot showed that the area $A(T)$ of a spherical triangle T, the radius R of the sphere on which it lies, and the interior angles β_i of T are related by the equation

$$\beta_1 + \beta_2 + \beta_3 - \pi = A(T)/R^2.$$

The left-hand side of this equation expresses the "angular excess" of the spherical triangle over the angle sum π for a plane triangle. A sphere has constant Gaussian curvature equal to $1/R^2$, and the second term in (5) vanishes, since the projection of a great circle onto the tangent plane is a line with curvature zero. Using the facts that at each vertex the interior and the exterior angles sum to π and that the Euler characteristic of T is one, we find that Harriot's result is precisely the Gauss-Bonnet theorem for spherical triangles.

Around 1630, Descartes generalized the Eucledean angle sum theorem from polygons to polyhedra. His result was that for any convex polyhedron the total

exterior solid angle is 4π, where at each vertex the solid exterior angle is by definition the sector of the unit ball bounded by planes normal to the edges. This result was lost until 1860, so it is only in hindsight that we can see it as a forerunner of the Gauss-Bonnet theorem. The surface of the polyhedron is closed (without boundary), so the first two terms in (5) do not appear. To make the connection with the Gauss-Bonnet formula, it is necessary to view the polyhedron as the limit of smooth surfaces. If curvature is correctly named, it is then reasonable that the third term in (5) approaches the sum of the solid angles, and Gauss proved this to be true. Since the Euler characteristic of a polyhedron can be computed to be two, we again have a special case of the Gauss-Bonnet theorem.

Let us now consider Stillwell's book itself. It offers a guided historical tour of mathematics from the time of Pythagoras to that of Gödel with separate chapters on subjects such as elliptic functions, differential geometry, and projective geometry, which are totally absent from most modern curricula. Also present are discussions of standard historical topics such as Greek geometry and the origins of calculus, topology, and group theory. The chapters are arranged in rough chronological order, and each concludes with a brief biography of one or two of the great mathematicians who have been most active in developing the topic discussed. Interspersing the mathematics with biographical sketches lightens the flow of ideas and reminds the reader that mathematics is created by interesting human beings. The exposition is masterfully linked together so that certain major themes such as curves and the solution of equations reappear in several chapters as the topic is attacked with different tools. Fundamental results such as Markov's theorem asserting the unsolvability of the homeomorphism problem for four manifolds, Bezout's theorem counting the intersections of algebraic curves, and Hermite's solution of the quintic using elliptic functions are discussed with some background. There are also statements of unsolved problems such as the Riemann hypothesis and the Poincaré conjecture that are part of basic mathematical culture and which ought to be mentioned in the undergraduate training. The discussion is at a deep enough level that I suspect most trained mathematicians will find much that they do not know as well as good intuitive explanations of familiar facts. The careful exposition, lightness of touch, and the absence of technicalities should make the book accessible to most senior undergraduate mathematics majors.

The selection of topics in a book of this sort is, of course, very much a matter of personal taste. Stillwell has chosen to say little about probability, statistics, or discrete mathematics, and applied mathematics is represented only by mechanics (in a very nice chapter which touches on Fourier series and the calculus of variations). I would have liked also to see a discussion of the origins of the Central Limit Theorem and of graph theory. The twentieth century is represented only by set theory, logic, and computation. A brief treatment of Lie groups and how they are used in modern physics would have presented a more balanced picture of contemporary mathematics.

The book contains interesting exercises often on topics which would enliven standard courses, for example, continued fractions, tesselations, and the pretty series expansion

$$\int_0^1 x^x \, dx = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \cdots.$$

(If you are unable to verify this formula, see page 187.) The bibliography is excellent. It is possible, for instance, to begin a more thorough study of cubics (as I have done) by following up the references cited.

There are of course slips and imprecisions in a book which strives to be a readable and nontechnical introduction to a very broad subject. In particular, there is an incorrect statement of the Riemann hypothesis ignoring the existence of the trivial zeros at the negative even integers (p. 129) and an assertion of the convergence of integrals in the Lebesgue theory with no hypothesis whatsoever (p. 318). Overall, however, the accuracy is good and the fluency of the writing and cogency of the arguments are extremely high. I recommend this book both to the professional for personal enjoyment and for use as a text to impart a large amount of mathematical culture to advanced students. If you are seeking a text for a capstone course, consider this book.

REFERENCES

1. W. W. Rouse Ball, On Newton's classification of cubic curves, *Proc. London Math. Soc.*, 22 (1890) 104–143.
2. Egbert Brieskorn and Horst, Knörrer, Plane Algebraic Curves, Birkhäuser Verlag, Basel, Boston, Stuttgart 1986.
3. Arthur Cayley, On the classification of cubic curves, *Trans. Camb. Phil. Soc.* vol. xi, part 1 (1866) 81–128.
4. Isaac Newton, The Mathematical Papers of Isaac Newton, ed. D. T. Whiteside, vol. 7, Cambridge University Press, Cambridge, 1976.
5. _____, The Mathematical Works of Isaac Newton, vol. 2, ed. D. T. Whiteside, Johnson Reprint Corporation, New York and London, 1967.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P*. *Moving Beyond Myths: Revitalizing Undergraduate Mathematics.* National Research Council. National Academy Pr, 1991, ix + 65 pp, (P). [ISBN: 0-309-04489-8] The long-awaited "MS 2000" report containing summary analysis and recommendations from the National Research Council's Committee on Mathematical Sciences in the Year 2000. Challenges U.S. colleges and universities to achieve the same stature for mathematics education as has been achieved during the past four decades in mathematical research. LAS

Elementary, T(13: 1). *Geometry for College Students, Second Edition.* Peter B. Geltner, Darrell J. Peterson. PWS-Kent, 1991, xiii + 336 pp, $29. [ISBN: 0-534-92472-7] Precise coverage of elementary concepts of traditional Euclidean geometry for college students with no geometry background. This edition has two-color format, rewritten definitions, more examples and exercises, and a new section on transformations, but retains two-column proofs. JNC

Education, P, L. *Developments in School Mathematics Education Around the World, Volume 2.* Eds: Izaak Wirszup, Robert Streit. NCTM, 1990, xiv + 475 pp, $20 (P). [ISBN: 0-87353-302-X] 25 papers presented in April 1988 at the Second International Conference on Mathematics Education sponsored by UCSMP, the University of Chicago's School Mathematics Project, supplemented by other invited reports. In three sections: international reports on the process of curriculum reform, research in mathematics education, and UCSMP reports. LAS

Education, P*. *A Call For Change: Recommendations for the Mathematical Preparation of Teachers of Mathematics.* Ed: James R.C. Leitzel. MAA Report. MAA, 1991, xiv + 47 pp, (P). [ISBN: 0-88385-072-9] Standards for undergraduate mathematical experiences for teachers in K-4, 5-8, and 9-12, building on six standards appropriate for all mathematics teachers. Not organized by courses but by "experiences" requiring at least 9 semester hours for K-4, 15 semester hours for 5-8, and a full mathematics major for 9-12. Written in parallel with the just-released NCTM *Professional Standards for Teaching Mathematics* (TR, May 1991). Emphasizes need to improve how college mathematics is taught in order that prospective teachers will learn by example how they can meet the NCTM *Standards* for school teaching. LAS

Education, P*. *Counting On You: Actions Supporting Mathematics Teaching Standards.* National Research Council, 1991, 17 pp, (P). A brief policy synopsis from MSEB that highlights changes recommended in the two NCTM *Standards* documents, in the new MAA recommendation on teacher education (*A Call for Change*), and in previous MSEB reports. Intended to help mobilize public and corporate support for improvements in mathematics teaching. LAS

History, P. *Konzepte des mathematisch Unendlichen im 19. Jahrhundert.* Ed: Gert König. Vandenhoeck and Ruprecht, 1990, 273 pp, (P). [ISBN: 3-525-40312-7] Revised texts of ten papers read at a symposium at the University of the Ruhr in June 1986. On the infinitely large and small in analysis, not Cantor's theory. JD-B

History, S, P, L.** *The Unreal Life of Oscar Zariski.* Carol Parikh. Academic Pr, 1991, xvii + 264 pp, $32.95. [ISBN: 0-12-545030-3] A compelling account of Zariski's extraordinary life (complementing his "real life" which is recorded in his scientific papers) based on interviews with family, colleagues, and students, and on tape-recorded interviews Zariski completed a few years before his death in 1986. Introduced by a foreword for non-mathematicians by Mumford; illustrated throughout with family photographs; concludes with a mathematical survey of Zariski's mathematical work (by Artin, *et al.*) and a complete Zariski bibliography. Award-winning author Parikh brings to life the intellectual ferment and international dimensions of Zariski's peripatetic career. LAS

Logic, T*(15-17: 1), L*. *Introduction to Modern Set Theory.* Judith Roitman. Wiley, 1990, xiii + 156 pp, $39.95. [ISBN: 0-471-63519-7] Well-written text provides introduction to modern and classical set theory. Includes models of set theory, consis-

tency results, large cardinals. No knowledge of logic assumed. KES

Logic, T(15-17), P, L. *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers.* Piergiorgio Odifreddi. Stud. in Logic & Found. of Math., V. 125. North-Holland (US Distr: Elsevier Science), 1989, xvii + 668 pp, $110.50. [ISBN: 0-444-87295-7] Text and reference manual covers recursive and partial recursive functions, recursively enumerable sets, hierarchies, Turing degrees, and m-degrees. Includes applications to logic, computer science, and other areas of mathematics. Note price. KES

Foundations, S(16-18), P, L. *Creative Understanding: Philosophical Reflections on Physics.* Roberto Torretti. Univ of Chicago Pr, 1990, xvi + 369 pp, $32.50 (P); $79.95. [ISBN: 0-228-80835-1; 0-226-80834-3] Einstein noted that concepts and fundamental laws of physics are not derived from experience, for they are "free inventions of the human mind." The aim of this book is to elucidate the workings of that creative understanding. Divided into five chapters labelled Observations, Concepts, Theories, Probability, and Necessity. The writing is compelling; the work makes a significant contribution to the philosophy of science. LCL

Foundations, P. *Higher Recursion Theory.* Gerald E. Sacks. Perspectives in Math. Logic. Springer-Verlag, 1990, xv + 344 pp, $89. [ISBN: 0-387-19305-7] Extensions of classical recursion theory (CRT), including hyperarithmetic sets (lifting CRT to least admissible set after ω), metarecursion (CRT on the recursive ordinals), α-recursion (CRT on the admissible ordinals), and E-recursion (extension to sets of arbitrary rank). RM

Foundations, T*(13-14: 1), S*, L*. *How To Read and Do Proofs: An Introduction to Mathematical Thought Processes, Second Edition.* Daniel Solow. Wiley, 1990, xx + 242 pp, $19.95 (P). [ISBN: 0-471-51004-1] Revised edition includes new chapter on nested quantifiers and more exercises. Excellent text for introduction to proof courses or supplement for other courses involving proof writing. (*First Edition,* TR, December 1982.) KES

Combinatorics, P. *Combinatorial Games.* Eds: Richard K. Guy, *et al.* Proc. of Symposia in Appl. Math., V. 43. AMS, 1991, xii + 233 pp, $52. [ISBN: 0-8218-0166-X] Ten papers from an AMS short course, including analysis of Gò and connections with coding theory, graph theory, number theory, and complexity theory. Contains unsolved problems and a selected bibliography. JPH

Combinatorics, T(16-17: 2), S. *Combinatorial Designs: Construction Methods.* Ian Anderson. Math. & Its Applic. Ellis Horwood (US Distr: Prentice Hall), 1990, 290 pp, $79.95. [ISBN: 0-7458-0349-0] Treats block designs, orthogonal Latin squares, Steiner systems, and tournament designs. Draws on research results through 1990. Suitable for seniors or graduate students. Prerequisite: linear algebra. AD

Discrete Mathematics, T(13-14: 1, 2). *Discrete Algorithmic Mathematics.* Stephen B. Maurer, Anthony Ralston. Addison-Wesley, 1991, xix + 894 pp, $38. [ISBN: 0-201-15585-0] Algorithmic approach but programming not a prerequisite. Possible use as a year-long alternative to calculus for well-prepared freshmen. Covers induction, graphs, counting methods, difference equations, probability, logic, sequences, series, and generating functions. Extensive chapter on basics and on linear algebra. AD

Linear Algebra, T(17-18: 1, 2), P, L. *Topics in Matrix Analysis.* Roger A. Horn, Charles R. Johnson. Cambridge Univ Pr, 1991, viii + 607 pp, $59.50. [ISBN: 0-521-30587-X] A sequel to *Matrix Analysis* by the same authors (TR, January 1987). This volume contains a more specialized set of topics: the field of values, matrix stability and inertia, singular value inequalities, matrix equations and Kronecker products, Hadamard products, and matrix functions. Designed to be useful as a reference and as a graduate-level textbook. AO

Linear Algebra, T(16-17: 2), L. *An Introduction to Linear Algebra.* L. Mirsky. Dover, 1990, viii + 440 pp, $10.95 (P). [ISBN: 0-486-66434-1] Abstract, rigorous approach to linear algebra. In addition to studying vector spaces, linear operators, and determinants, text covers canonical forms, orthogonal and unitary matrices, and quadratic forms. LC

Linear Algebra, S(14), C. *MAX: The MAtriX Algebra Calculator.* (IBM PC Software.) Charles H. Jepsen, Eugene A. Herman. Brooks/Cole, 1988; *Linear Algebra Problems for Computer Solution,* v + 144 pp, (P). [ISBN: 0-534-09300-0] A straightforward package for performing matrix calculations encountered in the typical first course in linear algebra. The accompanying text contains sample problems paralleling the standard course syllabus, mostly based on 160 selected matrices that come as part of the package. (New matrices can be entered from the keyboard or from a file.) LAS

Group Theory, S(18), P. *Lecture Notes in Mathematics-1456: Groups—Canberra 1989.* Ed: L.G. Kovács. Springer-Verlag, 1990, xii + 198 pp, $25 (P). [ISBN: 0-387-53475-X] Collection of papers presented during the Australian National University Group Theory Program in 1989, mostly during the Third International Conference on Group Theory and Related Topics in Canberra. Contains eight survey articles and eight research papers. List of participants. JS

Group Theory, S(18), P. *Discrete Subgroups of Semisimple Lie Groups.* G.A. Margulis. Ser. of Mod. Surveys in Math., Band 17. Springer-Verlag, 1991, ix + 388 pp, $85. [ISBN: 0-387-12179-X] Deals mainly with structure and classification problems for discrete subgroups of Lie groups, concentrating on lattices, which are defined as discrete subgroups of finite covolume. Topics include density and ergodicity, factor groups, boundary theory, rigidity, and arithmeticity. Appendices, historical notes, references, index. JS

Algebra, T*(16-18), S*, P*, L*. *Algebra, Vol-*

umes I-II. B.L. van der Waerden. Transl: John R. Schulenberger. Springer-Verlag, 1991, $29.95 each. *Volume I,* xiv + 265 pp, [ISBN: 0-387-97424-5]; *Volume II,* xii + 284 pp. [ISBN: 0-387-97425-3] This classic work, a reprint of the *Seventh Edition,* still sets the standard for the discipline, and offers so much of the best mathematics that it will never become unfashionable. *Volume I* contains the fundamentals of sets, groups, rings, ideals, and fields, followed by Galois theory and the theory of commutative fields. *Volume II* continues with valuation theory, group representations, theory of algebraic functions, and topological algebra. LCL

Algebra, T(17-18: 2), P, L. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions.* Bruce E. Sagan. Wadsworth, 1991, xv + 197 pp, $49.95. [ISBN: 0-534-15540-5] Aims to bring together many of the important results and techniques in the study of representations of the symmetric group. Except for some knowledge of linear algebra and elementary group theory, the book is relatively self-contained. It begins with an introduction to group representations. The second chapter studies representation of S_n, approaching it from the viewpoint of Specht modules. A combinatorial approach is used in Chapter 3. Included are the Frame-Robinson-Thrall hook formula, the Frobenius-Young determinantal formula, the Robinson-Schensted-Knuth algorithm, and Schützenberger's jeu de taquin. Finally, in Chapter 4, the symmetric functions, including the Schur functions, are introduced. The Jacobi-Trudi determinants, giving the Schur functions in terms of the elementary and complete symmetric functions, is proved. The book ends with the Littlewood-Richardson rule and the Murnaghan-Nakayama rule. LC

Algebra, P. *Cohen-Macaulay Modules over Cohen-Macaulay Rings.* Yuji Yoshino. London Math. Soc. Lect. Note Ser., V. 146. Cambridge Univ Pr, 1990, 177 pp, $24.95 (P). [ISBN: 0-521-35694-6] Revised version of lectures given at Tokyo Metropolitan University in 1987. Book addresses the following question: How can we classify CM modules over a given CM ring, or classify CM rings which have essentially a finite number of CM modules? The author presents a systematic, largely self-contained treatment of the subject. LCL

Algebra, P. *Lecture Notes in Mathematics-1448: Non-Commutative Ring Theory.* Eds: S.K. Jain, S.R. López-Permouth. Springer-Verlag, 1990, v + 166 pp, $19 (P). [ISBN: 0-387-53164-5] The proceedings (sixteen papers) of the Midwest Conference held at Ohio University in 1989. The intention for the conference was to emphasize and revitalize the interaction between researchers throughout the ring theoretic spectrum. LCL

Algebra, P. *The Polynomial Identities and Invariants of $n \times n$ Matrices.* Edward Formanek. CBMS Reg. Conf. Ser. in Math., No. 78. AMS, 1991, v + 57 pp, $22 (P). [ISBN: 0-8218-0730-7] Results in theory of polynomial identities related to the ring of $n \times n$ matrices over a field or commutative ring. Included are Kaplansky's Theorem, Posner's Theorem, and Artin's Theorem. LC

Calculus, T(13: 1-4). *Calculus, One and Several Variables.* Robert Ellis, Denny Gulick. Harcourt Brace Jovanovich, 1991, xiv + 1121 pp, $29.95. [ISBN: 0-15-505692-1] A 'black-and-white' version of the *Fourth Edition* of their calculus text; the idea being to offer an inexpensive text ($29.95!). It does, however, differ from the previous edition in two ways: applications of the integral precedes techniques for integration, and a chapter on complex numbers and functions is included. (*First Edition,* TR, August-September 1978; *Second Edition,* TR, October 1986; *Third Edition,* TR, June-July 1988; *Fourth Edition,* TR, November 1990.) KS

Calculus, S(13), C. *Calculus Calculator.* David Meredith. Prentice Hall, 1991, xii + 205 pp, $15 (P); $25 book and disk. [ISBN: 0-13-117441-X] An electronic blackboard on an IBM PC disk (inside book jacket). Two windows show input and output, the latter saved in a stack for subsequent use. Simple commands produce graphs in two and three dimensions. Includes capabilities for complex numbers and vector analysis. One chapter in accompanying manual is devoted to programming and development of scripts for instructional use. Manual contains dozens of applications, both of the calculator and programming features. Well-designed user interface is easy to learn and use. Requires 512K RAM. LAS

Complex Analysis, P. *Several Complex Variables VI: Complex Manifolds.* Eds: W. Barth, R. Narasimhan. Encylop. of Math. Sci., V. 69. Springer-Verlag, 1990, 310 pp, $59. [ISBN: 0-387-52788-5] Five longer survey papers, in expository style, on aspects of complex function theory and complex spaces: proper holomorphic mappings, Hodge theory, holomorphic transformation groups, surfaces, Teichmüller spaces. Each contains an extensive bibliography and a historical review. Published in honor of R. Remmert's sixtieth birthday. PZ

Complex Analysis, P.** *An Introduction to Multicomplex Spaces and Functions.* G. Baley Price. Pure & Appl. Math., V. 140. Marcel Dekker, 1991, xiii + 402 pp. [ISBN: 0-8247-8345-X] Multicomplex spaces are higher dimensional generalizations of the complex numbers, Banach algebras in their own right, obtained inductively by appropriately adjoining additional roots of -1. Such spaces support a theory analogous (in some respects, not in others) to classical complex function theory. Here the theory of the "bicomplex" space (with real dimension 4) is studied in detail; generalizations to higher dimensions are treated more briefly in a final chapter. Exposition is admirable: clear, readable, rich with exercises and useful asides to the reader. PZ

Differential Equations, S, C. *Phaseplane, The Dynamical Systems Tool, Version 3.0.* Bard Ermentrout. Brooks/Cole, 1990, xi + 131 pp, (P). [ISBN: 0-534-12894-7] An IBM-PC software pack-

age for numerical solution and trajectory display of ordinary differential and difference equations. Employs several methods for numerical integration (Euler, Runge-Kutta, Adams, Gear) and for calculation of equilibria. Menu-oriented screen display provides graphs with numerous options. Manual provides an overview of the various commands followed by ten tutorial examples based on built-in equations. Reference section provides details on file structures. LAS

Differential Equations, T*(14-15: 1), S, L*. *Differential Equations: A Dynamical Systems Approach, Part I: Ordinary Differential Equations.* John H. Hubbard, Beverly H. West. Texts in Appl. Math., V. 5. Springer-Verlag, 1991, xix + 348 pp, $39. [ISBN: 0-387-97286-2] A vivid, visual introduction to differential equations based less on analytic technique than on computer-assisted graphs to understand functions defined by differential equations. Makes effective use of descriptive concepts like "funnel," "fence," and "antifunnel" to identify important behavioral features. Chapters treat qualitative, analytic, numerical, and iterative methods as well as existence and uniqueness theorems. Part I comprises first five of seventeen chapters. Although intended for use with *MacMath,* the software package developed by the authors for this course, this text can be used with any system of similar capabilities. LAS

Partial Differential Equations, P. *Asymptotic Distribution of Eigenvalues of Differential Operators.* Serge Levendorskiĭ. Math. & Its Applic., V. 53. Kluwer Academic, 1990, xviii + 279 pp, $112. [ISBN:0-7923-0539-6] Using the Weyl-Hörmander calculus of pseudodifferential operators as a principal tool, this book describes ways to compute and analyze the asymptotics of eigenvalues of wide classes of differential operators. A book for researchers; not a text, and not for novices. AWR

Partial Differential Equations, T(18), S, P, L. *Variational Methods: Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems.* Michael Struwe. Springer-Verlag, 1990, xi + 244 pp, $39.50. [ISBN: 0-387-52022-8] Carefully written state-of-the-art overview. Begins with classical direct method and some recent extensions. Followed by chapter on minimax methods, and closes with chapter emphasizing limiting cases of the Palais-Smale condition. Preface provides nice brief history. Extensive bibliography. KS

Numerical Analysis, S(17). *Mathematical Analysis and Numerical Methods for Science and Technology, Volume 4: Integral Equations and Numerical Methods.* Robert Dautray, Jacques-Louis Lions. Transl: John C. Amson. Springer-Verlag, 1990, x + 465 pp, $89. [ISBN: 0-387-50209-2] The fourth in a six-volume series, contributed by well-known experts. Treats integral equations, mixed problems and Tricomi equations; comprehensive coverage of finite element methods for stationary boundary value problems. Studies briefly numerical methods for integral equations. LS

Numerical Analysis, P. *Lecture Notes in Mathematics-1457: Preconditioned Conjugate Gradient Methods.* Eds: O. Axelsson, L. Yu. Kolotilina. Springer Verlag, 1990, 196 pp, $25 (P). [ISBN: 0-387-53515-2] Eleven papers given at a 1989 conference at the University of Nijmegen, The Netherlands, are collected here in a book uncompromisingly directed to experts in the field. No effort in the preface or in any article to give an overview of the field, a context for the articles, or an indication of how they relate to each other. AWR

Numerical Analysis, T(2). *Numerical Methods for Conservation Laws.* Randall J. LeVeque. Lect. in Math. Birkhauser Boston, 1990, 214 pp, $24.50 (P). [ISBN: 3-8176-2464-3] Lecture notes for a mathematics course on conservation laws. Part I deals with theory of structure and solutions of the corresponding nonlinear systems. Part II deals with numerical methods, with emphasis on general tools rather than specific methods in detail. LS

Operator Theory, T(18), S, P. *Unbounded Operator Algebras and Representation Theory.* Konrad Schmüdgen. Operator Theory: Adv. & Applic., V. 37. Birkhauser Boston, 1990, 380 pp, $79.50. [ISBN: 0-8176-2321-3] The graph topology, topologies of O^*-algebras, the generalized Calkin algebra, commutants of O^*-algebras and *-representations, completely strongly positive mappings and the decomposition theory of states and *-representations are studied in detail. KS

Functional Analysis, S(18), P. *Lecture Notes in Mathematics-1449: Minimal Projections in Banach Spaces.* Włodzimierz Odyniec, Grzegorz Lewicki. Springer-Verlag, 1990, viii + 168 pp, $19 (P). [ISBN: 0-387-53197-1] Mainly devoted to the solution of uniqueness problems for minimal projections in Banach spaces. Also exhibits relevance of this problem to the uniqueness problem in mathematical programming and to the problem of characterization of Hilbert spaces. KS

Analysis, P. *Lecture Notes in Mathematics-1453: Global Analysis—Studies and Applications IV.* Eds: Yu. G. Borisovich, Yu. E. Gliklikh. Springer Verlag, 1990, 320 pp, $35 (P). [ISBN: 0-387-53407-5] Translation from Russian of nineteen articles from 1988-1989 in global analysis and nonlinear equations and in topological and geometric methods of analysis. Also includes a tribute to Kolmogorov. OJ

Analysis, P. *Representation of Lie Groups and Related Topics.* Eds: A.M. Vershik, D.P. Zhelobenko. Adv. Stud. in Contemp. Math., V. 7. Gordon and Breach, 1990, xiv + 557 pp, $294. [ISBN: 2-88124-678-8] Collection of articles on recent work in the Soviet Union on representations of "big" groups—ones that act like infinite dimensional Lie groups. Includes work on current groups (groups of sections of fiberings with a locally compact fibre), infinite linear groups ("classical" groups of operators on Hilbert spaces), and automorphism groups of manifolds. RM

Analysis, P. *Lecture Notes in Mathematics-1452: Number-Theoretic Analysis.* Eds: H. Hlawka, R.F. Tichy. Springer-Verlag, 1990, v + 220 pp, $24

(P). [ISBN: 0-387-53408-3] Continuation of seminar proceedings appearing in volumes *1114* and *1262*. KS

Algebraic Geometry, P. *The Curves Seminar at Queen's, Volume VII.* Anthony V. Geramita. Papers in Pure & Appl. Math., No. 85. Queen's U, 1990, 307 pp, (P). Contains two expository articles by Geramita (The non-singular cubic surface in $P^3 - II$) and by Robbiano (Hilbert functions), and nine research papers. SG

Algebraic Geometry, P. *Degeneration of Abelian Varieties.* Gerd Faltings, Ching-Li Chai. Ser. of Modern Surveys in Math., Band 22. Springer-Verlag, 1990, xii + 316 pp, $39.80. [ISBN: 0-387-52015-5] Study of degenerations (given by maps $G \to S$, S an irreducible scheme, G a group variety with abelian variety as generic fibre), and applications to constructions of moduli spaces, diophantine geometry. RM

Algebraic Geometry, P. *Lecture Notes in Mathematics-1454: p-adic Analysis.* Eds: F. Baldassarri, S. Bosch, B. Dwork. Springer-Verlag, 1990, v + 382 pp, $41 (P). [ISBN: 0-387-53477-6] A collection of twenty-plus papers touching on such topics as Drinfeld modules, rigid analytic spaces, zeta functions, approximations, and others. SG

Differential Geometry, T?(18), P. *Minimal Surfaces, Stratified Multivarifolds, and the Plateau Problem.* Dào Trong Thi, A.T. Fomenko. Transl. of Math. Mono., V. 84. AMS, 1991, ix + 404 pp, $152. [ISBN: 0-8218-4536-5] Good overview of current approaches and results for Plateau's problem. Presents solutions by Fomenko in the class of spectra of manifolds with fixed boundary and by Dào Trong Thi in the homotopy class of multivarifolds with a given boundary. Designed for advanced graduate students and researchers. Assumes basic knowledge in differential geometry, differential equations, algebraic topology, and the calculus of variations. The introduction on the one-dimensional Plateau problem and the first chapter on the history of Plateau's problem are accessible to a wider audience. Extensive bibliography. OJ

Geometry, S*(7-13). *Similarity.* Tom M. Apostol. Project Mathematics! California Institute of Technology (Distr: MAA), 1990, 28 minute videotape. *Program Guide and Workbook,* 30 pp, (P). One of a series of motivational and instructional videotapes intended primarily for students in grades 7-9. Production is consistent with normal TV fare: animation, stock footage, voice overlays, background music, etc. Focuses on scale models (moon lander) and scaling (maps) as an ubiquitous application of similarity—a "great triumph of Euclidean geometry." Uses computer graphics with gimmicks (ray guns for enlargement and reduction) to illustrate ratios, perimeters, areas, and volumes. Divided into "slates" coded with small on-screen numbers that are coordinated with the accompanying *Workbook.* Tapes are available from state Departments of Education, from the MAA, and from the Cal Tech bookstore. LAS

Topology, T(15-17), S, L. *Elementary Topology, Second Edition.* Michael C. Gemignani. Dover, 1990, xi + 270 pp, $6.95 (P). [ISBN: 0-486-66522-4] An unabridged and corrected republication of the 1972 *Second Edition* of the work originally published by Addison-Wesley in 1967 (TR, January 1968; Extended Review, January 1969; 1972 edition, TR, October 1973). An introductory text in which abstract concepts are introduced within the concrete context from which they developed (e.g., geometry, analysis, etc.). Topics include fundamentals of metric spaces, topologies, convergence, compactness, connectedness, homotopy theory, paracompactness, and complete regularity. A good buy. LCL

Operations Research, T(17-18: 1, 2), P. *Discrete Location Theory.* Eds: Pitu B. Mirchandani, Richard L. Francis. Wiley, 1990, xiv + 555 pp, $64.95. [ISBN: 0-471-89233-5] A collection of twelve articles written by distinguished researchers. Provides an overview of the theory and its application in real-world discrete location problems. The articles discuss the p-median problem, the uncapacitated facility location problem, the multi-period capacitated location problem, decomposition methods, covering problems, the quadratic assignment problem, competitive location problems, and the location of mobile units in a stochastic environment. Exercises and bibliographic references are included at the end of each chapter. AO

Optimization, S(17), P, L. *Introduction to Minimax.* V.F. Dem'yanov, V.N. Malozemov. Transl: D. Louvish. Dover, 1990, vii + 307 pp, $8.95 (P). [ISBN: 0-486-66423-6] This translation nicely pulls together in one place topics commonly spread out in American schools over approximation theory, the calculus of several variables, game theory, and linear programming. Chapter 1 treats best polynomial approximations (in the Chebyshev norm) in the discrete case, Chapter 2 in the continuous case. Chapter 3 discusses the linear minimax problem without constraints, and the same problems (still the discrete case) with convex constraints are taken up in Chapters 4 and 5. Chapter 6 concludes the matter by taking up the continuous case. The price, coupled with the quality of printing, makes this book a real bargain. (1974 Halsted Press edition, TR, March 1975.) AWR

Optimization, T*(15-17), L. *Introduction to Mathematical Programming: Applications and Algorithms.* Wayne L. Winston. PWS-Kent, 1991, xi + 794 pp. [ISBN: 0-534-92520-0] Aimed at undergraduate majors in business and operations research as well as MBA students, this text aims at being mathematically sound but not demanding in terms of preparation: no calculus is required until Chapter 12; the basic linear algebra needed is presented in Chapter 2. The use and interpretation of the output of LINDO and GINO are included, but set in separate sections that can be skipped by those who wish to do so. An attractive entry into this market. AWR

Optimization, P. *Parametric Optimization: Sin-*

gularities, Pathfollowing and Jumps. J. Guddat, F. Guerra Vazquez, H. Th. Jongen. Wiley, 1990, viii + 191 pp, $85. [ISBN: 0-471-92807-0] The subject matter is optimization problems in which some or all the individual data depend on just one parameter. The main tools for a solution algorithm are path-following methods, but these are shown not to be successful in every case, which leads to the subject of jumps in the set of local minimizers. The authors report on work centered at Kepler University (Linz) and Humboldt University (Berlin). Note the price per page. AWR

Dynamical Systems, S*(12-15). *Chaos, Fractals, and Dynamical Systems.* Robert Devaney. Science Television (Box 2498, Times Square Station, New York, NY 10108), 1989. 63 minute videotape, $59. [ISBN: 1-878310-00-3] A collage of materials on dynamical systems and Julia sets woven together by a lecture (a "talking head" in the upper right corner of the screen) that begins with generalities (weather, planets, stock markets) and very elementary exercises, but moves quickly to relatively sophisticated exploration of dynamics in the plane. Incorporates several of Devaney's short films on Julia sets of various sine and cosine functions in the complex plane. Illustrates various types of critical behavior both with static visuals (slide-like graphs) and with evolving Julia set films. Although the learning slope from beginning to end is rather steep, the film can serve to motivate and illustrate dynamical systems at many different levels. LAS

Modelling, S(15-17), P, L. *The Art of Modeling Dynamic Systems: Forecasting for Chaos, Randomness, and Determinism.* Foster Morrison. Wiley, 1991, xvii + 387 pp, $49.95. [ISBN: 0-471-52004-7] A non-technical survey, or field guide, of the methods of mathematical modelling with expository comments to indicate what each can do and cannot do. Techniques are illustrated with examples that can be read by anyone with an undergraduate education in science or mathematics. The purpose is to help the reader use and appreciate new developments in modelling techniques, including their limitations. LCL

Control Theory, P. *Lecture Notes in Control and Information Sciences-146: Minimum Entropy H_∞ Control.* D. Mustafa, K. Glover. Springer-Verlag, 1990, ix + 144 pp, $30 (P). [ISBN: 0-387-52947-0] An investigation of the entropy minimization problem. The approach taken here exploits recent progress in the state-space solution to the H_∞ control problem and explicit formulae can be obtained for the solution. LCL

Control Theory, P. *Lecture Notes in Mathematics-1459: Optimal Control of Nonsmooth Distributed Parameter Systems.* Dan Tiba. Springer-Verlag, 1990, vii + 159 pp, $20 (P). [ISBN: 0-387-53524-1] Overview and recent research on distributed control problems governed by nonlinear evolution equations (parabolic and hyperbolic) in absence of differentiability. Includes nonlinear delay differential equations, boundary control, elliptical problems, and optimal design. RM

Control Theory, P. *Distributed Parameter Control Systems: New Trends and Applications.* Eds: Goong Chen, *et al.* Lect. Notes in Pure & Appl. Math., V. 128. Marcel Dekker, 1991, x + 516 pp, (P). [ISBN: 0-8247-8444-8] A collection of seventeen papers based on invited presentations at a workshop held August 4-11, 1989 at the University of Minnesota. AO

Probability, T(16-18), L. *Probability for Applications.* Paul E. Pfeiffer. Texts in Stat. Springer-Verlag, 1990, xix + 679 pp, $59.95. [ISBN: 0-387-97138-6] An introduction to advanced probability theory aimed at teaching theory while understanding basic examples. Chock full of understandable, illustrative examples and exercises. Topics are standard: probability measures, set theory, independence, conditional probability, random variables and distributions, expectation, Radon-Nikodym and conditional expectation, stochastic processes. MK

Probability, P. *Lecture Notes in Mathematics-1427: École d'Été de Probabilités de Saint-Flour XVIII—1988.* A. Ancona, D. Geman, N. Ikeda. Springer-Verlag, 1990, vii + 330 pp, $35 (P). [ISBN: 0-387-53508-X] Lecture notes from three courses given at l'École d'Ètè de Calcul des Probabilités de Saint-Flour, August 21-September 7, 1988. Ancona: "Théorie du potentiel sur les graphes et les variétés." Geman: "Random fields and inverse problems in imaging." Ikeda: "Probabilistic methods in the study of asymptotics." LC

Stochastic Processes, S(18), P. *Introduction to Stochastic Integration, Second Edition.* K.L. Chung, R.J. Williams. Prob. & Its Applic. Birkhauser Boston, 1990, xv + 276 pp, $34.50. [ISBN: 0-8176-3386-3] Presents a modern version of the theory of stochastic integration, in particular integration with respect to a local martingale with continuous paths. This edition includes discussion of stochastic differential equations, Cameron-Martin-Girsanov formula, the Feynman-Kac-Schrödinger development, and reflected Brownian motions. Proofs are carefully detailed, but motivation and exposition is left for the reader to discover. (*First Edition,* TR, December 1984.) MK

Stochastic Processes, S(18), P. *Empirical Processes: Theory and Applications.* David Pollard. NSF-CBMS Reg. Conf. Ser. in Prob. & Stat., V. 2. Institute of Math Statistics, 1990, viii + 86 pp, (P). [ISBN: 0-940600-16-1] A monograph based on lectures given at the University of Iowa in July 1988. Results concern sums of independent stochastic processes allowing generalizations to nonidentically distributed random variables. Makes use of Vapnik-Červonenkis classes to generalize and extend Glivenko-Cantelli and Donsker's theorems for empirical distribution functions, as well as empirical central limit theorems, uniform laws of large numbers, least absolute deviation estimators, estimation for censored data, and more. MK

Stochastic Processes, T(18). *Continuous Martingales and Brownian Motion.* Daniel Revuz, Marc

Yor. Grund. der math. Wissenschaften, B. 293. Springer Verlag, 1991, ix + 533 pp, $89. [ISBN: 0-387-52167-4] Develops theory for continuous Martingales, Markov processes, and Gaussian processes by focusing on the theory of Brownian motion. Important applications include stochastic calculus and excursion theory. Lots of exercises ranging from straightforward to very difficult. MK

Elementary Statistics, S(13-16), L. *How to Tell the Liars from the Statisticians.* Robert Hooke. Marcel Dekker, 1983, xv + 173 pp. [ISBN: 0-8247-1817-8] Nice collection of brief accounts showing that statistical reasoning affects much of our lives—from drug testing to mass production. Well-written for those with little or no mathematical or statistical background. Good supplementary text in introductory-level courses. MK

Statistics, P. *Theory of Random Determinants.* V.L. Girko. Math. & Its Applic., V. 45. Kluwer Academic, 1990, xxv + 677 pp, $245. [ISBN: 0-7923-0233-8] An encyclopedic coverage, expanded by 50% from the original Russian edition that dates from 1980, including original results which have not been published previously. The theory of determinants of random matrices (random determinants) involves the theory of probability and statistics and its applications. Note price! LCL

Statistics, T(17), S, P. *Conditioning Diagnostics: Collinearity and Weak Data in Regression.* David A. Belsley. Prob. & Math. Stat. Wiley, 1991, xx + 396 pp, $59.95. [ISBN: 0-471-52889-7] Exposes state-of-the-art regression diagnostic methodology. Updates and revises material from *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* Belsley, Kuh, and Welsch (TR, March 1981). Includes new data sets, extensions to nonlinear models, and simultaneous-equation regression. Topics include collinearity; weak data; ill-conditioning diagnostics; geometry of mean centering, correlation, and least squares; condition indices; variance-decomposition proportions; interpreting collinearity diagnostics with applications; and discussion of influential observations. Also indicates open research problems. MK

Statistics, S(13). *Calculator Enhancement for Introductory Statistics, Preliminary Edition.* Clemson University Calculator Project. Eds: Iris Fetta, D.R. LaTorre. Harcourt Brace Jovanovich, 1991, x + 160 pp, $6 (P). [ISBN: 0-15-505678-6] A collection of applications of the programmable, graphing calculator (in this case, the Sharp EL-5200) to problems and methods of elementary probability and statistics: e.g., histograms, simulation, various distributions, sampling, hypothesis testing, linear regression, etc. Not a textbook or study guide: application of technology, not mathematical content, is emphasized. Although tied tightly to the Sharp (e.g., programs are displayed as keystroke sequences), the ideas contained are readily transferable to other machines. Diligent use of the supplement seems, as claimed, likely to "enhance" understanding, both of technology and of mathematics. PZ

Elementary Computer Science, T(13-14). *Advanced LOGO: A Language for Learning.* Michael Friendly. Lawrence Erlbaum Assoc, 1988, xiv + 661 pp, $24.95 (P); $79.95. [ISBN: 0-89859-933-4] Intended for high school and college students with some experience in LOGO. After a quick review of basic LOGO, topics such as procedures, lists, recursion, trees, graphics, and dynamics are treated in depth. SG

Programming, T(14-16: 1), L. *Object-Oriented Database Programming.* Suad Alagić. Texts & Mono. in Comput. Sci. Springer-Verlag, 1989, xv + 320 pp, $49.50. [ISBN: 0-387-96754-0] A text on programming methodology and associated concepts, techniques, and language notation devoted to database applications, using an extension of Modula-2. "Modern programming methodology is just as important" in application-oriented problems as in system programming or mathematics-oriented algorithms. Presumes an introductory programming course, basic mathematics. Exercises. RB

Algorithms, P. *Uses of Randomness in Algorithms and Protocols.* Joe Kilian. ACM Distinguished Dissertations. MIT Pr, 1990, 235 pp, $39.95. [ISBN: 0-262-11153-5] The author's MIT dissertation in which randomness is used to attack three problems: finding a probabilistic algorithm to test primality; zero knowledge proofs about committed bits; and proof systems for determining if a string is in a language. A lively and well-written book. SG

Computer Systems, S(16-17), P. *OSI Explained: End-to-End Computer Communication Standards, Second Edition.* John Henshall, Sandy Shaw. Ser. in Comput. & Their Applic. Ellis Horwood (US Distr: Prentice Hall), 1990, 249 pp. [ISBN: 0-13-639451-5] Relatively accessible explanation of Open Systems Interconnection standards and procedures for computer networks, with special emphasis on the top four levels (application, presentation, session, transport) which specify how networked systems cooperate to support complex applications. RM

Computer Systems, T(14-15: 1), S, L. *A Programmer's Guide to File Processing.* Douglas L. Cashing. PWS-Kent, 1991, xv + 169 pp, $28. [ISBN: 0-534-92567-7] Language independent, conceptual treatment of file structures and processing, suitable for ACM Curriculum '78 CS 5 course. Progresses from non-technical overview through implementation issues, including overview of database management and sorting. RM

Computer Systems, P. *Advanced Topics in Data-Flow Computing.* Eds: Jean-Luc Gaudiot, Lubomir Bic. Prentice-Hall, 1991, xx + 626 pp. [ISBN: 0-13-006503-X] A monograph based on a workshop in Eliat, Israel associated with the 1989 Sixteenth International Symposium on Computer Architecture. Surveys of the principles and 20-year evolution of data-flow computer architectures; 21 research papers on architecture issues, data-flow/von Neumann hybrids, high-level languages and compilation issues,

scheduling and resource management, special purpose systems. RB

Artificial Intelligence, S(16-17), P. *Artificial Intelligence: A Tool for Industry and Management.* Hans W. Gottinger, H. Peter Weimann. Ser. in Artif. Intellig. Ellis Horwood (US Distr: Prentice Hall), 1990, 158 pp. [ISBN: 0-13-048372-9] Overview, in the form of a technology assessment, of some major commercial developments and applications of artificial intelligence in areas such as expert systems, natural language processing and speech recognition, robotics, database systems, and decision support systems. RM

Computer Science, S?(16), P, L. *The Practice of Prolog.* Ed: Leon S. Sterling. MIT Pr, 1990, 312 pp, $37.50. [ISBN: 0-262-19301-9] Collection of articles covering fairly large (and interesting) applications developed in Prolog: database prototyping, VLSI editor, assisting register transfer design, parallelizing Pascal compiler, partial evaluation system, natural language explanation, learning, and stream data analysis. Good reading for a programmer who wants to see how to build real (i.e., not toy) systems in Prolog. RM

Computer Science, P. *Communication Complexity: A New Approach to Circuit Depth.* Mauricio Karchmer. ACM Doctoral Dissertation Award. MIT Pr, 1989, 69 pp, $25. [ISBN: 0-262-11143-8] This volume is the author's doctoral dissertation from MIT, which won the award for the outstanding Ph.D. dissertation of 1988. It addresses the problem of the circuit complexity of Boolean functions. The author develops a new approach for analyzing the complexity, called communicational complexity, which provides theoretical bounds on the complexity of computer circuits. GMS

Computer Science, S(16-17), L. *Readings in Object-Oriented Database Systems.* Eds: Stanley B. Zdonik, David Maier. Ser. in Data Management Systems. Morgan Kaufmann, 1990, x + 629 pp, $39.95 (P). [ISBN: 0-55860-000-0] Good selection of foundational and recent papers on object-oriented approach to database systems, including fundamentals of object orientation, semantic models, implementation issues, extensions to the relational model, and applications. RM

Computer Science, T(15-16), S. *The Craft of Prolog.* Richard A. O'Keefe. MIT Pr, 1990, xix + 387 pp, $34.95. [ISBN: 0-262-15039-5] Prolog is a programming language based on the principles of the first-order predicate logic. This text is an introduction to the language as well as a primer on the use of logic as a vehicle for problem solving and controlled deduction. Prolog has received a great deal of attention recently due to its ever-increasing use in the field of expert systems and artificial intelligence. GMS

Computer Science, T(15-16), S. *How to Write Parallel Programs: A First Course.* Nicholas Carriero, David Gelernter. MIT Pr, 1990, 232 pp, $29.95. [ISBN: 0-262-03171-X] Parallel processing, the technique of problem solving using multiple computers, is becoming an increasingly popular technique for attacking very, very large computations. The authors introduce the basic concept of parallel processing and show a number of solutions implemented using the C and Linda programming languages, and the popular MIMD parallel architecture. These languages and architecture are the most widely used and widely available. GMS

Computer Science, P, L. *Beauty Is Our Business: A Birthday Salute to Edsger W. Dijkstra.* Eds: W.H.J. Feijen, *et al.* Texts & Mono. in Comput. Sci. Springer-Verlag, 1990, xix + 453 pp, $49. [ISBN: 0-387-97299-4] How to honor the sixtieth birthday of a researcher and teacher who has profoundly impacted the intellectual foundations of computer science? This commemorative volume was prepared in secret by dozens of prominent colleagues he has influenced, and was presented to him in a surprise birthday conference! Fifty-four eclectic contributions seeking the best in elegance and style. RB

Computer Science, P. *Parallel Processing for Scientific Computing.* Eds: Jack Dongarra, *et al.* SIAM, 1990, 454 pp, $49.50 (P). [ISBN: 0-89871-262-9] Proceedings of the fourth SIAM conference on the topic, December 1989, in Chicago. Forty-nine papers on matrix computations, numerical methods, differential equations, massively parallel computation, performance, and tools. RB

Applications, T, S, P, L. *Fourier Methods for Mathematicians, Scientists and Engineers.* Mark Cartwright. Math. & Its Applic. Ellis Horwood (US Distr: Prentice Hall), 1990, 326 pp, $37.50. [ISBN: 0-13-327016-5] In two parts: Fourier Theory (220 pages) and Fourier Applications (88 pages). The first 100 pages on series and transforms contain no surprises and are very readable. The writing throughout is lean and clear. Chapters on convolution, filtering, correlation, and transforms follow a brief chapter on two-dimensions. Part I concludes with a chapter on the discrete Fourier transform and its uses. Part II is loaded with many unusual, but admittedly sketchy, applications in physics (quantum theory, lasers, scattering, diffraction, telescopes, interferometry, impedance), chemistry (spectroscopy), life science (vision, hearing, speech analysis), and earth science (water waves, turbulence, sliding glaciers, seismology). For upper-class students and beginning graduate students. Exercises at chapter ends in Part I with solutions to some. Excellent references, especially to topics in Part II. JDEK

Applications, P. *Computing Methods in Applied Sciences and Engineering.* Eds: Roland Glowinski, Alain Lichnewsky. SIAM, 1990, 446 pp, $49.50 (P). [ISBN: 0-89871-264-5] Proceedings of the Ninth International Conference on Computing Methods sponsored by INRIA (1990). Papers on state-of-the-art methods for solutions for large linear and nonlinear systems, optimization, dynamics and turbulence, computational chemistry and physics, parallel and supercomputing. RM

Applications (Biological Science), S(18), P,

L. *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity.* Eds: Monica G. Turner, Robert H. Gardner. Ecological Stud., V. 82. Springer-Verlag, 1991, xv + 536 pp, $98. [ISBN: 0-387-97326-5] Compiles a list of techniques used to study spatial heterogeneity. Essentially a collection of papers, each with an extensive bibliography, divided into four main sections: introduction, analysis and interpretation of patterns, model development and simulation, and synthesis. SM

Applications (Biological Science), T(16-17), P, L. *Community Food Webs: Data and Theory.* Joel E. Cohen, Frédéric Briand, Charles M. Newman. Biomath., V. 20. Springer-Verlag, 1990, xi + 308 pp, $79. [ISBN: 0-387-51129-6] Intended for community ecologists with a quantitative bent and teachers/researchers involved with random graphs. An application of combinatorics. Several unsolved mathematical problems are given. SM

Applications (Economics), P. *The Economic Design of Control Charts.* Elart von Collani. Skripten zur Math. Stochastik. BG Teubner Stuttgart, 1989, xii + 171 pp, DM 29 (P). [ISBN: 3-519-02729-1] Simple graphical methods are developed to determine the economic design of $\overline{x}-, c-$, and $p-$charts. RWJ

Applications (Fluid Dynamics), P, L. *Annual Review of Fluid Mechanics, Volume 23, 1991.* Eds: John L. Lumley, Milton Van Dyke, Helen L. Reed. Annual Reviews, 1991, x + 674 pp, $40. As usual with this series, this volume contains an interesting mix of theory and real-world applications. Eighteen articles ranging from "Theory of Hurricanes" to "Symmetry and Symmetry-Breaking Bifurcations in Fluid Dynamics." JO

Applications (Information Theory), T(18), P. *Introduction to Shannon Sampling and Interpolation Theory.* Robert J. Marks II. Texts in Electrical Engin. Springer-Verlag, 1991, xiii + 324 pp, $59. [ISBN: 0-387-97391-5] Study of generalizations of the "cardinal" series (bandlimited signal) of Shannon sampling theory, for sampling and subsequent restoration by interpolation of continuous signals and images. Analysis of effects of truncation, noise, jitter, on restoration error, generalizations to higher dimensions, optimal sampling strategies. RM

Applications (Information Theory), T(16-17: 1), S. *Information Theory.* Robert B. Ash. Dover, 1990, xi + 339 pp, $8.95 (P). [ISBN: 0-486-66521-6] Relatively self-contained introduction to the theory of classical Shannon information theory. Topics include analysis of channel models (with and without memory), coding systems and coding theorems, statistical properties of sources, noiseless coding, error-correcting codes, continuous channels. Exercises. Reprint of the 1965 Wiley edition. RM

Applications (Physical Science), P. *Modeling One Dimensional Pattern Formation by Anti-diffusion.* J.F. Kaashoek. CWI Tract, V. 72. Mathematisch Centrum, 1990, vi + 271 pp, Dfl 67 (P). [ISBN: 90-6196-389-3] Introduces the definition of pattern formation, providing physical backgrounds and examples. Presents models of phase transition, reaction-diffusion equations, mass conservation evolution equations, etc. Investigates theoretical properties of solutions. Reports numerical simulations for the models. LS

Applications (Physics), S(18), P. *Wavelets: Time-Frequency Methods and Phase Space, Second Edition.* Eds: J.M. Combes, A. Grossmann, Ph. Tchamitchian. Springer-Verlag, 1990, ix + 331 pp, $59 (P). [ISBN: 0-387-53014-2] Proceedings of an International Conference held in Marseille, France, December 14-18, 1987. MU

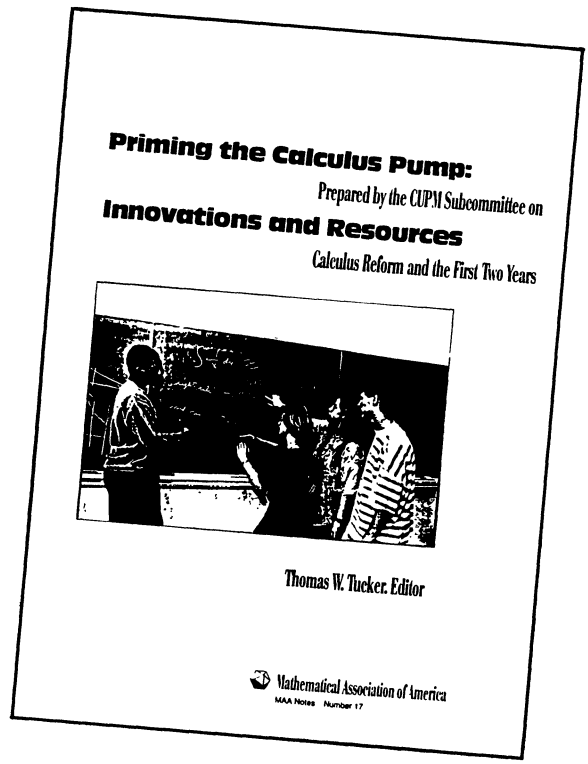
Applications (Physics), P, L*. *Selected Papers, S. Chandrasekhar, Volume 6: The Mathematical Theory of Black Holes and of Colliding Plane Waves.* S. Chandrasekhar. Univ of Chicago Pr, 1991, xix + 739 pp, $35 (P); $79.95. [ISBN: 0-226-10101-0; 0-226-10100-2] Final volume in these *Selected Papers.* Thirty-seven papers, mostly dating from the years 1975-85, in which Chandrasekhar returned to a subject he pioneered forty years earlier to develop "*ab initio*" the whole theory of black holes. Introduced by an extraordinarily warm and engaging personal retrospective on this phase of Chandra's career written by Chandra's colleague and frequent co-author Basilis Xanthopoulos, who was murdered in Crete on November 27, 1990 while giving a seminar lecture. (Chandra describes their association as "the most binding in all my years in science.") LAS

Applications (Physics), P. *Spectral and Scattering Theory for Wave Propagation in the Perturbed Stratified Media.* Ricardo Weder. Appl. Math. Sci., V. 87. Springer-Verlag, 1991, 188 pp, $39. [ISBN: 0-387-97357-5] A stratified medium is one whose physical properties depend only upon a single coordinate, whereas if perturbed, then locally the physical properties depend on all the coordinates. An understanding of wave propagation in these media is needed for areas such as ocean acoustics, integrated optics, and wave guides. This self-contained text includes many new results and methods published for the first time. MPR

Reviewers

RB: Richard Brown, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; AD: Amy Davidow, Macalester; SG: Steven Galovich, Carleton; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LS: Li Sheng, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

Calculus Reform



PRIMING THE CALCULUS PUMP: INNOVATIONS AND RESOURCES

Prepared by the CUPM Subcommittee on Calculus Reform and the First Two Years

Thomas W. Tucker, Editor

Among our many educational crimes, the one that bears the greatest responsibility for the shortage of young Americans choosing mathematics, physics or engineering as careers may very well be the wretched way we teach calculus.

Peter D. Lax

In the four years since the Tulane conference issued a call for a "lean and lively calculus," there have been countless articles, contributed paper sessions, talks, and panels. There have also been many individuals and institutions who have responded with energy and imagination in getting grants, setting up laboratories, developing software, assigning students projects, and rethinking from top to bottom what should go into a calculus course. They are changing not only what they teach, but also, what is far more difficult, how they teach. It is one thing to ask for reform in conferences and articles; it is another thing entirely to do it in the classroom, as these people are doing.

Unfortunately, these activities do not always get the publicity they deserve. That is the purpose of this book—to disseminate rapidly to the mathematical community detailed examples of calculus reform at work. The ten featured projects in this book, together with abstracts of more than sixty other projects and a collection of reference materials and resources, are designed to give individuals and departments a concrete idea of what they themselves can do, as well as information on how to do it, what to use, and who to contact.

328 pp., Paper, 1990
ISBN 0-88385-067-2

List: $20.00

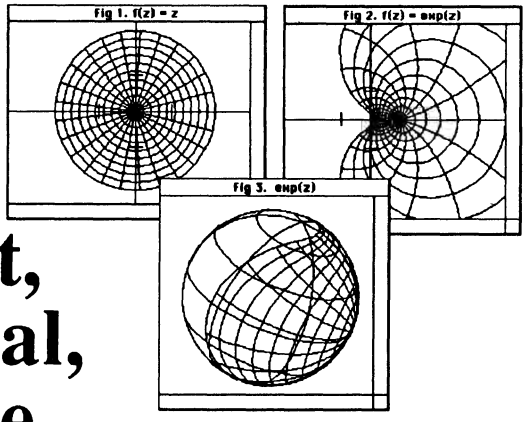
Catalog Number NTE-17

ORDER FROM



Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

Software and video tapes for the student, professional, and anyone who loves mathematics.



Ask for our latest catalog.

Lascaux Graphics 3220 Steuben Ave.
Bronx, NY 10467 USA (800) 338-0993

STUDIES IN THE HISTORY OF MATHEMATICS

STUDIES IN THE HISTORY OF MATHEMATICS

Esther R. Phillips, Editor

Esther Phillips has brought together a collection of articles showing the sweep of recent scholarship in the history of mathematics. The material covers a wide range of current research topics: algebraic number theory, geometry, topology, logic, the relationship between mathematics and computing, partial differential equations, and algebraic geometry.

320 pp., 1987, ISBN 0-88385-128-8

List: 36.50 MAA Member: $28.00

Catalog Number MAS-26

This is an excellent book! It is a very interesting and exciting book to read. The author does an extremely nice job of bringing together most, if not all, the mathematicians that were involved in a particular area of mathematics. The sources listed at the end of each section give the reader an opportunity to look up other resources pertaining to the particular subjects, a feature that is definitely lacking in many history books. The content of the book is choice. The professional mathematician would definitely want to have a copy of this book.

Barney Erikson in *The Mathematics Teacher*

Order from: The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

MATHEMATICS: Queen and Servant of Science

by E. T. Bell

"Bell is a lively, stimulating writer, inoffensively crotchety and opinionated, with a good sense of historical circumstance, a fine impatience with humbug, a sound grasp of the entire mathematical scene, and a gift for clear and orderly explanation."

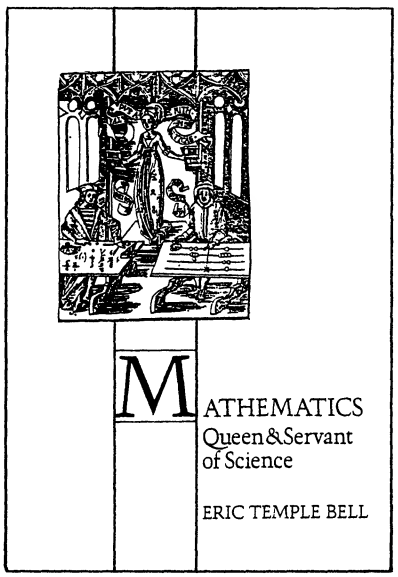
James R. Newman, ***The World of Mathematics***

"Bell's writing style makes you feel as if you are a fortunate student spending time with a charming and gifted teacher. Not a history of mathematics, but a wide ranging exposition of ideas in the subject, the book is a classic and should be required reading for every teacher of mathematics."

Harold Jacobs, Grant High School, Van Nuys, California

"This book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students of mathematics at all levels."

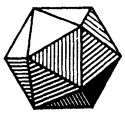
Morris Kline, Courant Institute



xv + 437 pp. Paperbound
Republished 1987.
ISBN-0-88385-446-3
List: $18.50
MAA Member: $14.00

This is an absorbing account of the development in pure and applied mathematics from the geometry of Euclid more than two-thousand years ago to some of the more recent developments in mathematical physics. Published 30 years ago, MATHEMATICS: QUEEN AND SERVANT OF SCIENCE remains a valuable tool for teachers, students, and all those enthusiastic about mathematics. The book includes 20 chapters on selected areas in pure and applied mathematics, including discussions on topics such as geometry, algebra, logic, probability, infinity, transformations, rings, groups, topology, and continuity. Also covered are subjects such as: Mersenne primes, quantum mechanics, and three-valued logic. The book is not a history of mathematics, but rather an enticing sampling of selected topics in mathematics and science.

The book contains fascinating information on the personalities and philosophies of the great mathematicians, plus perceptive resumes of their specific contributions to the moving force of this great science.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

NEW! VERSION 2

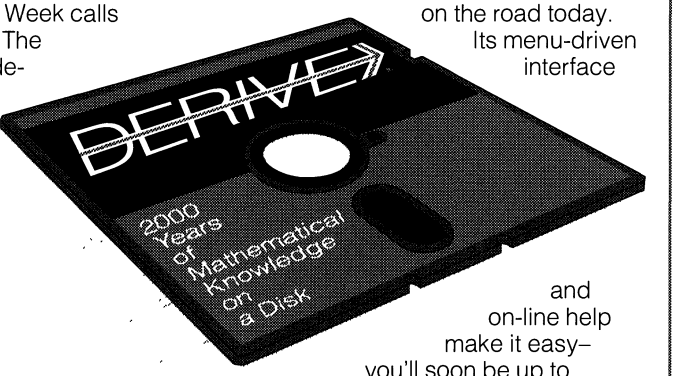
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE
May 29, 1990
Derive, Version 1 6

DERIVE®– A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math. instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

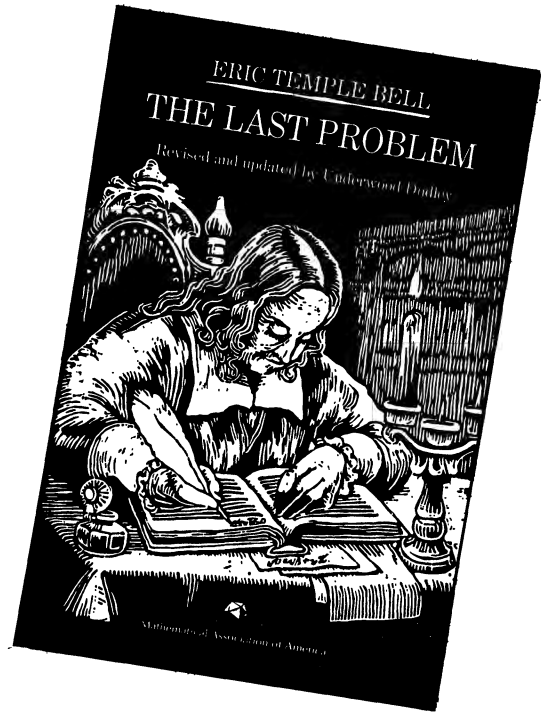
And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

Revised and Updated



THE LAST PROBLEM

E. T. Bell
Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990
ISBN-0-88385-451-1

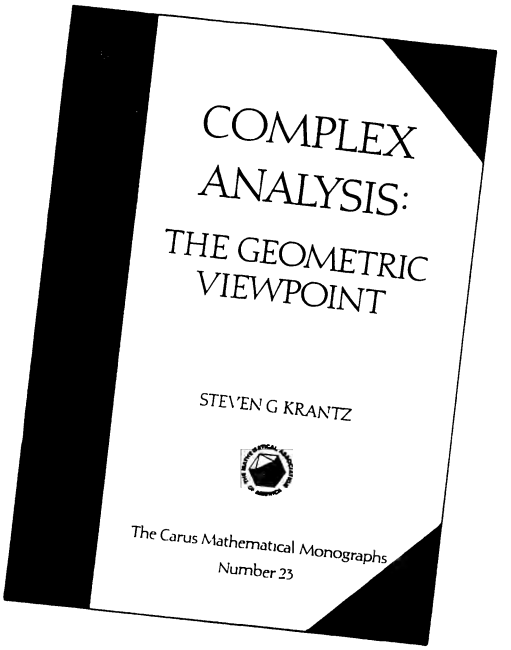
List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

New From Steven G. Krantz



COMPLEX ANALYSIS:
THE GEOMETRIC VIEWPOINT

Steven G. Krantz

Geometric methods have been used in complex analysis since the 1930s when Lars Ahlfors discovered that they give a nice way to look at the Schwarz lemma. Since that time they have become a central part of the research activities of complex analysis. However these important techniques have never found their way into a text accessible to a broad audience.

Steven G. Krantz, a leading worker in complex analysis and a well-known mathematical expositor, has written the first book explaining how complex analysis can be studied using methods of geometry. Assuming no background in Riemannian geometry, and only one semester of complex analysis, Krantz explains the role of Hermitian metrics and of curvature in understanding the Schwarz lemma, normal families, Picard's theorems, conformal mappings, and many other topics. A minimum of geometric formalism is used to gain a maximum of geometric and analytic insight. The climax of the book is an introduction to several complex variables from the geometric viewpoint. Poincaré's theorem, that the ball and bidisc are biholomorphically inequivalent, is discussed and proved.

Except for the minimal background requirements, the book is self-contained. A review of relevant topics in the classical theory of one complex variable is provided. The style is light and inviting. The book is a must for anyone with an interest in complex analysis. Take a glance at the main chapter headings and order your copy today

- Principal Ideas of Classical Function Theory,
- Basic Notions of Differential Geometry,
- Curvature and Applications,
- Some New Invariant Metrics,
- A Glimpse of Several Complex Variables

210 pp., 1990, Hardbound,
ISBN 0-88385-026-5

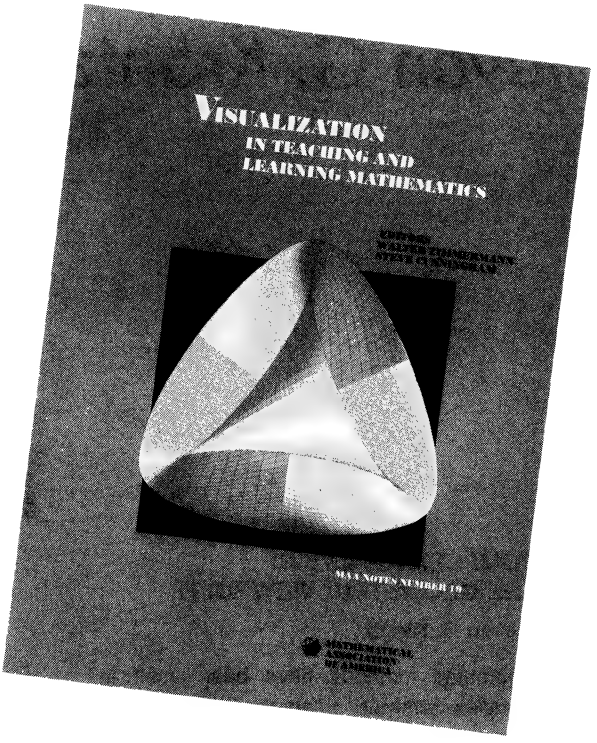
List $22.00 MAA Member $18.50

Catalog Number CAM-23

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

New from the MAA



VISUALIZATION IN TEACHING AND LEARNING MATHEMATICS

Walter Zimmermann
and Steve Cunningham, Editors

We are proud to announce the publication of "Visualization in Teaching and Learning Mathematics," a project of the MAA Committee on Computers in Mathematics Education. Included in this volume are several beautiful full-color figures that will enhance the understanding of the mathematical concepts presented. Both hand-drawn and computer generated graphical representations of mathematical concepts, principles or problems are used in the book.

The twenty papers in the book give an overview of research, analysis, practical experience, and informed opinion about the role of visualization in teaching and learning mathematics, especially at the undergraduate level. Visualization, in its broadest sense, is as old as mathematics, but progress in computer graphics has generated a renaissance of interest in visual representations and visual thinking in mathematics.

Six of the papers in this volume deal broadly with philosophical, pedagogical or technological issues in visualization. The remaining papers address the role of visualization in three-dimentional geometry, calculus, differential equations, linear algebra, complex analysis, numerical analysis, stochastic processes and statistics. Most of these papers discuss the role of interactive computer graphics in teaching and learning mathematical ideas.

230 pp., 1991, ISBN 0-88385-071-0

List: $20.00

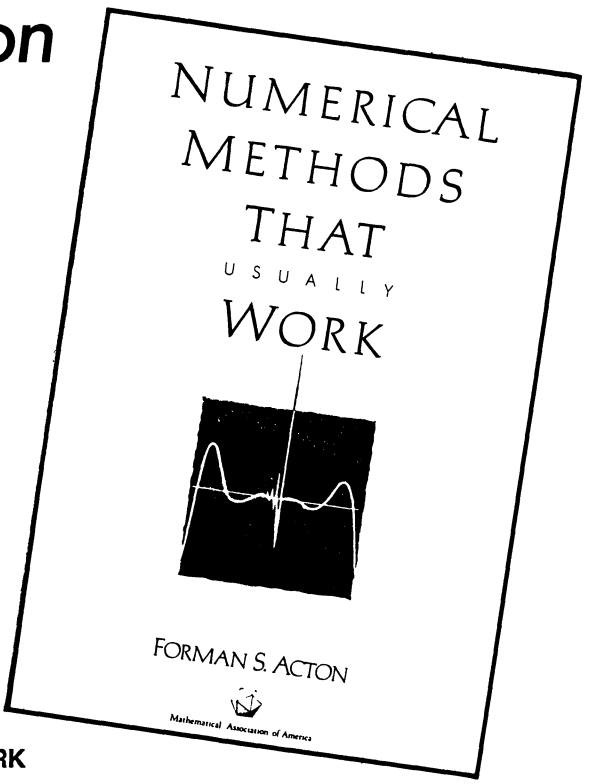
Catalog Number NTE-19

Order From



Mathematical Association of America
1529 18th Street, N.W.
Washington, D.C. 20036

New From Forman S. Acton



NUMERICAL METHODS THAT WORK

Forman Acton

Numerical Methods That Work, originally published in 1970, has been reissued by the MAA with a new preface and some additional problems. Acton deals with a common sense approach to numerical algorithms for the solution of equations: algebraic, transcendental, and differential. It assumes that a computer is available for performing the bulk of the arithmetic. The book is divided into two parts, either of which could form the basis of a one-semester course in numerical methods. Part I discusses most of the standard techniques; roots of transcendental equations, roots of polynomials, eigenvalues of symmetric matrices, and so on. Part II cuts across the basic tools, stressing such common problems as instabilities in extrapolation, removal of singularities, and loss of significant figures. The book is written with clarity and precision, intended for practical rather than theoretical use.

An eminently readable and very well movitivated introductory text.
Mathematics of Computation

A first-rate book which can be used either as a text or reference.
Choice

560 pp., Paperbound, 1990
ISBN 0-88385-450-3

List $16.50 MAA Member $13.00

Catalog Number NMTW

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

Introducing E.Z. Math for the HP 48SX

E.Z. Math is a program for the Hewlett Packard 48SX calculator written, conceived and programmed by Raymond La Barbera and the E.Z. Software Company. It comes on a 128K plug-in ROM card and with a 188 page manual. It is designed to be used by math and science students, teachers, parents and business people.

What Does E.Z. Math Do?

E.Z. Math has six modules dealing with graphs, loans, savings, numbers, games and music:

Graphs. E.Z. Math covers the entire high school and college graphing curriculum, from elementary algebra to advanced calculus. The student can choose from 188 families of equations, inequalities, functions, systems of equations and systems of inequalities, in rectangular, polar and parametric form, all laid out and arranged in an easy-to-use, logically organized system of menus. E.Z. Math makes the graphic analysis of polynomial, rational, trigonometric, hyperbolic, logarithmic and exponential functions quick and easy.

Loans. E.Z. Math makes it easy to solve problems involving fixed rate mortgages on houses, condos, co-ops, and other real estate investments. Anyone wanting to purchase a home can quickly compute the monthly mortgage payment, the amount that can be borrowed, the number of years needed to pay off the loan and the annual interest rate and see a detailed breakdown of each payment in a complete amortization table.

Savings. E.Z. Math makes it easy to solve problems involving a single deposit or repeated deposits to a savings account, certificate of deposit [C.D.], term deposit [T.D.], money market account or other such investment. Anyone interested in saving money can quickly and easily compute the amount accumulated, the amount that must be deposited, the number of years necessary to keep the money on deposit and the annual interest rate.

Numbers. E.Z. Math makes it easy to handle numbers of various types. Elementary and high school students will find it easy to compute all the factors and the prime factorization of a natural number. In addition, finding the greatest common factor, least common multiple and average of any set of natural numbers is a snap. Any set of rational numbers, whether in whole number, fraction or mixed number form, may be added, subtracted, multiplied or divided yielding an answer expressed as a fraction or mixed number in lowest terms or as an integer. Any set of complex numbers may be added, subtracted, multiplied, divided or raised to a power. Any number of terms of the sequences of perfect nth powers, triangular numbers, binomial coefficients, Fibonacci numbers and multiples of a number can be easily computed.

Recreation. E.Z. Math provides a couple of strategy number games and an introduction to the musical capabilities of the HP48SX. These are provided both for fun and as an encouragement to explore further some of the many features of the calculator, including a bit of the built-in rich programming language.

Education. Although using most features of E.Z. Math involves no more than plugging in the ROM card and following the various on-screen menus, the E.Z. Math User's Manual explains very clearly and precisely in complete detail all features and display screens of the program. Just as important, 46 pages of the manual are devoted to simple, intuitive explanations of the basic concepts involved in graphs, numbers, sets, variables, equations, savings and loans.

What's Special About E.Z. Math?

E.Z. Math has many special and unique features which contribute to its practicality and ease of use. Here are some of them:

- A very easy-to-use, logically organized user interface makes possible the full use of all E.Z. Math features without having to open the 850 page HP Owner's Manual.
- To reduce the possibility of user error, all answer screens repeat all entered data when displaying the answer .
- All menu screens, answer screens and graphics screens are easily printable on the HP82240 infrared printer.
- Careful error trapping takes care of invalid user input without crashing the program.
- Absolutely no user contact with the stack is required to use all E.Z. Math features. Yet, going to the stack without quitting E.Z. Math is just a keystroke away.
- Our easy quit feature restores all user system flags and custom menus, leaves no "garbage" on the stack and gets rid of variables which are no longer needed.
- Our easy off feature allows turning off the calculator without quitting the program.

How to Order E.Z. Math

E.Z. Math is available at a retail price of $125.00, plus $5.00 shipping and handling. We accept payment by check, money order, COD, VISA, Mastercard, American Express and school purchase orders. If within 30 days of receipt you find that E.Z. Math fails to fulfill your expectations, we'll gladly take back your copy for a prompt, courteous refund. To order copies of E.Z. Math or to obtain information about about ordering E.Z. Math bundled with HP 48SX calculators and classroom overhead projector kits, please write or call:

SMI Corporation
Dept EMM
250 West New Street, Suite 48
Kingsport, TN 37660
(800) 234-0123 or (615) 378-4821 or (615) 245-8982 (Fax)

Help your students discover more meaningful relationships.



Renewed for '91: a free classroom display device with purchase of 30 calculators.

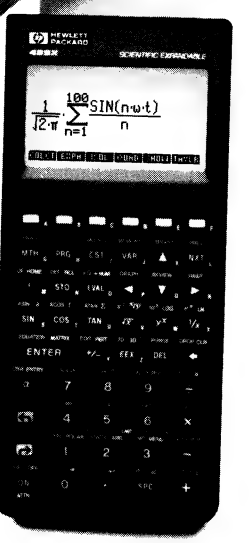
Showing is much more powerful than telling. So we've developed special classroom displays for our most advanced calculators.

The HP 48SX scientific expandable calculator, the cost-effective HP 28S and the new HP 48S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculators.

If your department or students purchase 30 HP 48SX, HP 28S or HP 48S calculators (or a mix of all three), we'll give you free an HP 48SX and plug-in classroom display (a $900 retail value). Or a classroom display version of the HP 28S (a $600 retail value). And free teaching materials.

Call **(503) 757-2004** from 8am to 3pm PDT for details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1991, and applies only to college and high school instructors.

HEWLETT PACKARD



© 1991 Hewlett-Packard Company PG12005

Springer for Mathematics

H.-O. Peitgen, H. Jürgens, D. Saupe, all University of Bremen, FRG, in cooperation with **E. Maletsky**, Montclair State College, NJ; **T. Perciante**, Wheaton College, IL; and **L. Yunker**, West Chicago Community High School, IL

Fractals for the Classroom

Introduction to Fractals and Chaos

Fractals for the Classroom breaks new ground as it brings an exciting branch of mathematics into the classroom. The book is a collection of independent chapters on the major concepts related to the science and mathematics of fractals. Written at the mathematical level of an advanced secondary student, **Fractals for the Classroom** includes many fascinating insights for the teacher and the individual who is interested in learning the historical background and the amazing potential of fractals, but who does not want to get tied down with too many technical terms. Illustrations from a wide variety of applications are integrated with an enjoyable text to help bring the concepts alive and make them understandable to the average reader.

(Published in cooperation with the National Council of Teachers of Mathematics).
1991/app. 450 pages, 125 illus./Hardcover/ $29.00/ISBN 0-387-97041-X

Also by **Peitgen, Jürgens, Saupe, Maletsky, Perciante**, and **Yunker**

Fractals for the Classroom

Strategic Activities, Volume 1

This first volume of strategic classroom activities is designed to develop, through a hands-on approach, a basic mathematical understanding and appreciation of fractals. These strategic activities have been developed from a sound instructional base, stressing the connections to the contemporary curriculum. A slide package containing some of the highest quality fractal images available is included.

(Published in cooperation with the National Council of Teachers of Mathematics).
1991/144 pages.,75 illus., 9 color slides/ Softcover/$19.95/ISBN 0-387-97346-X

H.-O. Peitgen, H. Jürgens, D. Saupe
Written by **T. Eberhardt** and **M. Parmet**

The Beauty of Fractals Lab

Graphics Software for the Macintosh Version 1.0

1990/$49.00/ISBN 0-387-14205-3

F.J. Flanigan, San Jose State University, CA; and **J.L. Kazdan**, University of Pennsylvania, PA

Calculus Two

Linear and Nonlinear Functions
Second Edition

With revisions by **D.L. Frank, B.E. Fristedt** and **L.F. Gray**

This book introduces linear algebra in an intuitive geometric setting as the study of linear maps and uses these simple linear functions to study more complicated linear functions. This second edition has been revised throughout and new material has been added including more information on linear algebra, Gaussian elimination, and dimension bases. Many exercises and illustrations have been included to make the text more usable for students.
1990/619 pp., 207 illus./Hardcover/$39.95
ISBN 0-387-97388-5
Undergraduate Texts in Mathematics

Three Easy Ways to Order:

- **Call**: Toll-Free **1-800-SPRINGE(R)**: 1-800-777-4643. In NJ call 201-348-4033 (8:30 AM — 4:30 PM EST). **Your reference number is S823.**
- **Write**: Send payment plus $2.50 for postage and handling to: Springer-Verlag New York, Inc., Dept. S823, PO Box 2485, Secaucus, NJ 07096-9812.
- **Visit**: Your local technical bookstore.

Instructors: Call or Write for information on textbook examination copies!



Springer-Verlag

New York • Berlin • Heidelberg • Vienna • London • Paris • Tokyo • Hong Kong • Barcelona

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 7 August–September 1991

Contents

(ISSN 0002-9890)

RTICLES

Number Theory as Gadfly B. MAZUR 593

OTES

3*Q*; 1*A* HENRY S. WARREN, JR. 611

An Occurrence of the Ballot Numbers in Operator Theory LAWRENCE H. RIDDLE 613

Derangements, Permanents, and Christmas Presents STEPHEN G. PENRICE 617

Distance Functions and Topologies FRED GALVIN AND S. D. SHORE 620

HE TEACHING OF MATHEMATICS

A Characterization of Baire Spaces DRAGAN JANKOVIĆ, MAXIMILLIAN GANSTER, AND IVAN REILLY 624

Which Latin Squares are Cayley Tables? MAN-KEUNG SIU 625

An Overview of the Calculus Curriculum Reform Effort: Issues for Learning, Teaching, and Curriculum Development JOAN FERRINI-MUNDY AND KAREN GEUTHER GRAHAM 627

An Early Introduction to Dynamics MARYSIA T. WEISS 635

A Simple Interpretation of the Complex Contour Integral ALAN GLUCHOFF 641

'ROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 645

Advanced Problems and Solutions 655

REVIEWS

Beyond the Third Dimension. By Thomas Banchoff H. S. M. COXETER 668

Global Differential Geometry. Edited by S. S. Chern THOMAS F. BANCHOFF 669

TELEGRAPHIC REVIEWS 672

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $33 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

Number Theory as Gadfly

B. MAZUR, *Harvard University*

DR. MAZUR received his Ph.D. from Princeton University in 1959, and was a Junior Fellow at Harvard University from 1961–64. Since then he has been at Harvard University with frequent visits to I.H.E.S. in France. He is a member of the U.S. National Academy of Sciences and has received the Veblen Prize (in geometry) and the Cole Prize (in number theory) from the AMS.



(This is an expository article which evolved from notes written in preparation for a 40-minute talk for a general audience at the "Symposium on Number Theory," held in Washington D.C. on May 4, 1989 under the auspices of the Board on Mathematical Sciences of the National Research Council. To make the text more informative the original version has been supplemented with lots of commentary, a section (§4) has been added which may be useful to readers familiar with the classical theory of modular forms, and an appendix has been added which is meant for an even more specialized audience. I am thankful to P. Diaconis, J. Mazur, K. Ribet and J.-P. Serre, who read early drafts of this paper, and whose suggestions were very helpful to me.)

1. Introduction. When a friend saw the title to my talk he asked if what I had in mind was the well-known fact that number theory has an annoying habit: the field produces, without effort, innumerable problems which have a sweet, innocent air about them, tempting flowers; and yet... the quests for the solutions of these problems have been known to lead to the creation (from nothing) of theories which spread their light on all of mathematics[1], have been known to goad mathematicians on to achieve major unifications of their science[2], have been known to entail painful exertion in other branches of mathematics to make those branches serviceable[3]. Number theory swarms with bugs, waiting to bite the tempted flower-lovers who, once bitten, are inspired to excesses of effort!

Well, perhaps that summarizes the general aim of my talk—but, to put it more gently, I want to spend a few minutes considering one example (a conjecture, in fact) which shows how Number Theory can sometimes contrive to be a helpful, and possibly inspirational, goad to the rest of the Mathematical Sciences.

The most celebrated of all deceptively simple (and still unsolved!) problems in Number Theory is surely *Fermat's Last Theorem*[4]. Its curious history (whose statement first occurs as a marginal commentary on the equation arising from the Pythagorean theorem) is so well known, it needn't be rehearsed here. Professional mathematicians, after Fermat, have approached Fermat's Last Theorem with

[1] e.g., Kummer's theory of ideals

[2] e.g., Grothendieck's theory of schemes

[3] e.g., The theory of group representations, and in particular, the "Langlands program"

[4] For a detailed account of the recent work on this see Oesterlé's Bourbaki report [**O**] listed in the References for §4.

mixed feelings. Kummer, for example, called it "more of a joke than a pinnacle of science," and he came as close as anyone has come to proving it! Gauss wouldn't work on it, even when urged by his friends to do so in order to get the cash reward which had been offered for its solution.

But Fermat's Last Theorem has always been the darling of the amateur mathematicians, and as things have progressed, it seems that they are right to be enamored of it: Despite the fact that it resists solution, it has inspired a prodigious amount of first-rate mathematics. Despite the fact that its truth hasn't a *single* direct application (even within number theory!) it has, nevertheless, an interesting *oblique* contribution to make to number theory: its truth would follow from some of the most vital and central conjectures in the field. Although others are to be found, Fermat's Last Theorem presents an unusually interesting "test" for these conjectures.

My aim is to describe, in some detail, one of these grand conjectures (due to Shimura, Taniyama and Weil[5]) which, even though still unproved, plays a structural and deeply influential role in much of our thinking and our expectations in Arithmetic. Thanks to recent work of Frey, Serre, and Ribet[6], it has a large number of applications, Fermat's Last Theorem among them. As I shall not have time to make clear, but hope, at least, to make believable, the conjecture of Shimura-Taniyama-Weil is a profoundly unifying conjecture—its very statement hints that we may have to look to diverse mathematical fields for insights or tools that might lead to its resolution[7].

As we shall see, the conjecture of Shimura-Taniyama-Weil would imply a strange and important connection between the elliptic curves that arise in Arithmetic (we'll get to that shortly!) and the *Hyperbolic Plane*.

I mentioned above that the conjecture of Shimura-Taniyama-Weil has as one of its consequences, Fermat's Last Theorem. As everyone knows, Fermat's Last Theorem is an assertion about the family of (Diophantine[8]) equations

$$X^N + Y^N = Z^N$$

[5]The fact that this conjecture has also been referred to as the *Weil conjecture*, the *Taniyama-Weil conjecture*, and the *Taniyama conjecture* points to the difficulty in assigning to it a clear attribution. It was originally formulated as a problem by Taniyama in a conference in 1955 and was published in Japanese, in Sûgaku 7 (1956) p. 269. A more precise formulation corresponding to the modern form of the conjecture—involving important information concerning the conductor—was implicitly suggested by subsequent work of Weil which had the effect of bringing the problem to the attention of a large audience. The most precise version of this conjecture to date, which brings in the crucial issue of fields of definition, incorporates work of Shimura, Eichler, and others (see footnote 15 below, and the technical appendix at the end). For a moving evocation of the life and times of Taniyama (as well as an English translation of the original statement of Taniyama's problems) see the article "Yutaka Taniyama and his time, very personal recollections," Goro Shimura, *Bull. London Math. Soc.* 21 (1989) 186–196.

[6]See [S], [Fr 1, Fr 2], and [R] listed in the references at the end of §4.

[7]It does not seem unnatural to look to differential geometry for progress with this conjecture, or to partial differential equations and the study of the eigenvalue problem for elliptic operators, or to the representation theory of reductive groups.... It would be no surprise if ideas from the classical theory of one complex variable and the Mellin transform were relevant, or of Algebraic Geometry.... But perhaps one should also look in the direction of Kac-Moody algebras, loop groups, or $\mathscr{D}$-modules, perhaps to ideas that have been, or will be, imported from Physics....

[8]The adjective "Diophantine" is in honor of the Alexandrian mathematician Diophantos (perhaps A.D. 250) and signals vaguely a type of equation not dissimilar from those Diophantos considered.

for $N = 3, 4, 5, \ldots$, or with little loss of generality, for odd prime exponents N. But despite the elegance, and evident symmetries of the above family of equations, there is no denying that, after all, it is merely *one* family of Diophantine equations. What is so excellent about these particular equations?

Now, the urge to put any *single* instance in an appropriate general context before dealing with it mathematically is strong. Consider, for example, the way François Viète, the modern inventor of algebra, expressed that urge at the end of his treatise (c. 1591) by saying that algebra appropriated for itself "the proud problem of problems: which is to *leave no problem unsolved.*" More to the point, consider the most celebrated of recent Diophantine results, valid in a *truly* general context: the theorem of Faltings (conjectured originally by Mordell) which asserts that any algebraic equation in two variables, and of genus[9] greater than or equal to 2, has only a finite number of rational solutions.

But what *is* an appropriate general context in which to place the Fermat family of Diophantine equations, and what *is* the appropriate Diophantine question to ask? Despite the fact that the Fermat problem has been with us for three centuries I don't believe that we have any thoroughly comfortable answer, even to this modest question. One may always take recourse (in cases where it is not clear how to "correctly" generalize a problem) in the reliable method of *kicking the problem a bit*, to get a "nearby" one.... A relatively conservative move in this direction, in the case of Fermat's equation, is to allow general coefficients in the equation, say one coefficient for starters—for example, fix a nonzero integer A, and consider the family

$$A \cdot X^N + Y^N = Z^N,$$

and then ask: Is there an exponent N_0 such that for exponents N (or for prime exponents N) greater than N_0 there is no triple of integers (X, Y, Z), none zero, solving the above equation? In this slight perturbation of Fermat's original problem a few minutes of reflection will convince one to be circumspect in framing precise conjectures... (e.g., consider $A = 2$). Nevertheless the conjecture of Shimura-Taniyama-Weil has an impressive power of prediction concerning the

About Diophantos' personal history little is known, save what can be gleaned from the following problem which occurs in a collection, the *Palatine Anthology*, compiled, scholars believe, no more than a century after his death:

> Here you see the tomb containing the remains of Diophantos, it is remarkable: artfully it tells the measures of his life. The sixth part of his life God granted him for his youth. After a twelfth more his cheeks were bearded. After an additional seventh he kindled the light of marriage, and in the fifth year he accepted a son. Elas, a dear but unfortunate child, half of his father he was and this was also the span a cruel fate granted it. He consoled his grief in the remaining four years of his life. By this devise of numbers, tell us the extent of his life.

[9]The *genus* g of an algebraic curve is a nonnegative integer which was originally introduced by Riemann and defined by "topological means." It also has an "algebraic" definition, and as such is an intrinsic invariant of the field of algebraic functions on the curve. If the curve is given as the locus of zeroes of a homogeneous form of degree d in three variables in the projective plane then $g \leqslant (d-1)(d-2)/2$ with equality holding if and only if the curve has no singularities. In contrast to the genus, however, the *degree* is not given by the field of functions of the curve alone: it is defined in terms of the representation of the curve in projective space. For this reason it is more natural to look to the *genus* rather than to the *degree as* an invariant which determines "Diophantine behavior."

nature of nontrivial integer solutions for these families. For example,[10] using work of Frey, Serre, and Ribet, the Shimura-Taniyama-Weil conjecture would imply that the above equation has no such solutions for prime exponents $N > 7$ if A is any power of 3, or of 5, or of 7 (or of 11, 13, 17, 19, 23, 29, 53 or 59, for that matter[11], provided that N doesn't divide A), and it would guarantee the existence of an N_0 such that for prime exponents $N \geqslant N_0$ there are no such solutions if A is any power of any odd prime not of the form $2^n \pm 1$ (i.e., if A is neither a power of a Mersenne nor of a Fermat prime).

In summary, the conjecture of Shimura-Taniyama-Weil seems to be getting into the thorny thicket of these Diophantine issues—seems to be giving reasons why some (but not all!) of these equations cannot have solutions—seems to be beginning to put such Diophantine problems in a "context."

It also relates them to the extraordinary geometric questions to which we shall now turn.

2. Euclidean and non-Euclidean covering mappings. One of the mysteries of the Shimura-Taniyama-Weil conjecture, and its constellation of equivalent paraphrases, is that although it is undeniably a conjecture "about arithmetic," it can be phrased variously, so that: in one of its guises, one thinks of it as being also deeply "about" integral transforms in the theory of one complex variable; in another as being also "about" geometry[12].

The more striking of these two formulations is the geometric one. To explain it we need to review a few basics of geometry: *symmetries*, *orbits*, *orbit spaces*, *covering mappings*, and the interesting concept of "*uniformization*" We'll build things up slowly by first considering these notions in a relatively simple context (on the *Real Line*), and then in the two contexts (*Euclidean* and *Non-Euclidean*) necessary for the actual "geometric" statement of the conjecture.

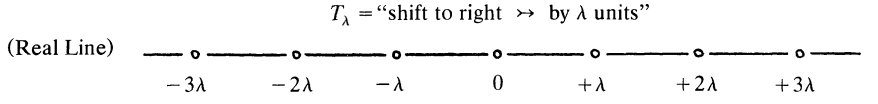


FIG. 1. The "lattice" $\Lambda = \{0, \pm\lambda, \pm 2\lambda, \ldots\}$ is the orbit of 0 under the translation T_λ.

(I). On the real line.

Let λ be a positive real number and let T_λ denote the operation on the real line $\mathbb{R}$ consisting of "shifting all points in $\mathbb{R}$ to the right" by λ (synonymously: *translation* by λ; in equations: $T_\lambda(x) = x + \lambda$). We view T_λ as a *symmetry* of $\mathbb{R}$. The iterated translates of 0 by T_λ and by its inverse $T_{-\lambda}$ give us a discrete evenly spaced "lattice" Λ in $\mathbb{R}$ consisting of all integral multiples of λ.

[10]See thme. 2 and subsequent remarks in section 4.3 of J.-P. Serre's [S] listed in the references at the end of §4.

[11]And with more work, one could surely produce more such consequences of Shimura-Taniyama-Weil.

[12]The equivalence of these two formulations is due to Weil, following upon work of Hecke.

The *orbit* of a real number x with respect to the lattice Λ, is the collection of real numbers obtained from x by translating x by iterates of T_λ and by its inverse $T_{-\lambda}$ (that is, the orbit is the collection of real numbers $x + N \cdot \lambda$ where N runs through all integers). The lattice Λ itself is the orbit containing the origin. Given a fixed real number A, any orbit for Λ has a unique representative in the interval $A \leqslant x < A + \lambda$.

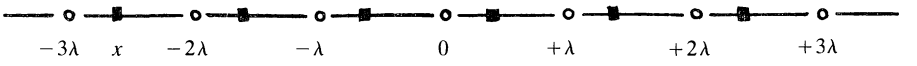


FIG. 2. The black squares mark the "orbit" of some general real number x.

The *orbit space* of $\mathbb{R}$ with respect to the symmetry T_λ (or one might also say: with respect to the lattice Λ) is simply the collection of orbits (with respect to Λ). Call the orbit space $\mathbb{R}/\Lambda$. One can think of the orbit space as obtained from $\mathbb{R}$ by "identifying" any point x in $\mathbb{R}$ with any iterated translate of x by T_λ and $T_{-\lambda}$, i.e., with all points $y \in \mathbb{R}$ such that $y = x + N\lambda$ for some integer N. The easiest way to visualize this is by wrapping the real line (viewed as a helix in the figure below) around a circle (say a unit circle in the complex plane) in such a manner that points on $\mathbb{R}$ of distance λ apart map to the sample point of the circle:

$$z \mapsto e^{2\pi i z/\lambda}. \qquad (*)$$

The circle, then, may be taken to be the orbit space. The displayed mapping $(*)$ can be called a *covering mapping* and has the property that the pre-image with respect to $(*)$ of any point on the circle consists in precisely one orbit with respect to Λ.

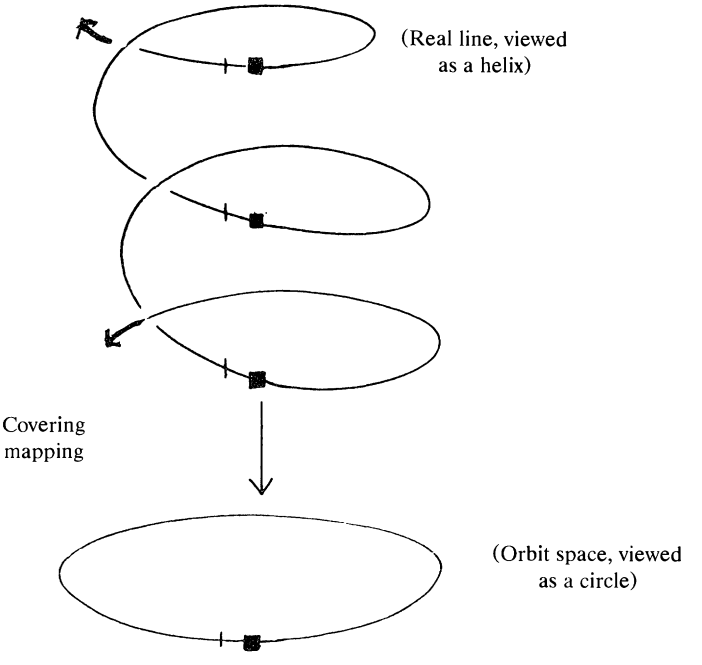


FIG. 3

and the various symmetries obtainable by iterations of them ($x \mapsto N_1 \cdot \lambda_1 + N_2 \cdot \lambda_2$, for integers N_1 and N_2). Here λ_1, λ_2 are two complex numbers such that 0, λ_1, and λ_2 are not collinear (and therefore "generate" a lattice Λ as in FIGURE 5).

An orbit with respect to Λ (that is, the set of points in the complex plane which can be obtained by the iterated application of the translations $T_{\lambda_1}, T_{\lambda_2}$ and their inverses to a single point in the complex plane) is simply a "displaced lattice" as in FIGURE 6 below. Any such orbit has exactly one representative in the (half-closed) parallelogram as in the next figure, whose vertices are 0, λ_1, λ_2, and $\lambda_1 + \lambda_2$.

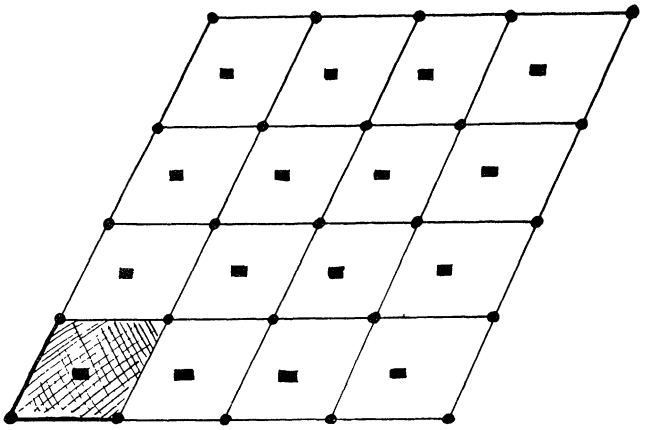


FIG. 6. The ■'s mark an orbit with respect to the lattice.

As in the example of $\mathbb{R}$ given previously, by the orbit space with respect to the two symmetries T_{λ_1} and T_{λ_2} (or, we may say, with respect to the lattice Λ) we mean *the collection of orbits* with respect to Λ. We may visualize this orbit space as the parallelogram with vertices $0, \lambda_1, \lambda_2$, and $\lambda_1 + \lambda_2$ "wrapped up" as shown in FIGURE 7 below. Topologically it is a "torus," i.e., the surface of a doughnut. Call

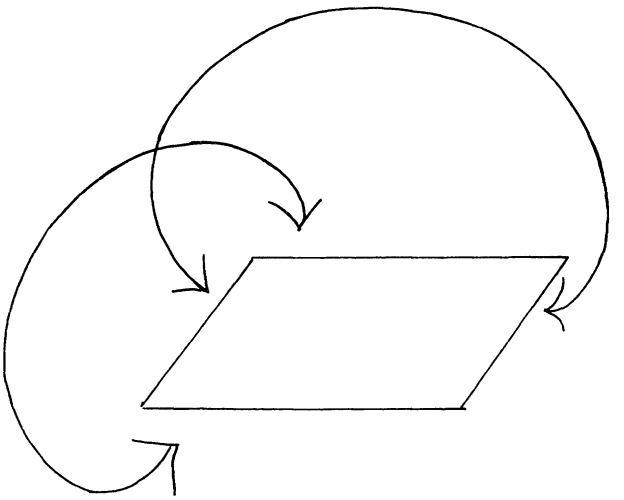


FIG. 7

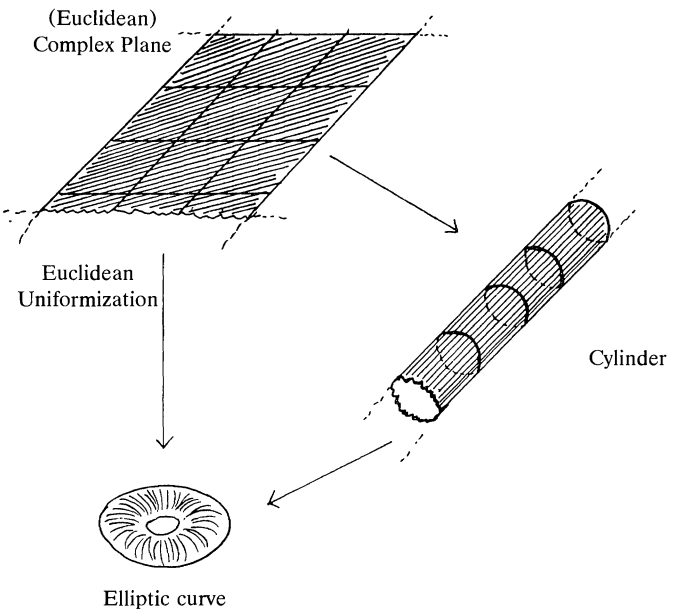


FIG. 8. The "covering mapping" which brings the complex plane $\mathbb{C}$ to the orbit space $\mathbb{C}/\Lambda$ may also be visualized as a two-stage process, where in the first stage the plane is wrapped around a cylinder, and in the second stage the cylinder is wrapped around a torus.

the orbit space $\mathbb{C}/\Lambda$. The mapping which sends each point in the complex plane to the orbit which contains it is our *covering mapping* $\mathbb{C} \to \mathbb{C}/\Lambda$:

We wish to think of the orbit space $\mathbb{C}/\Lambda$ as inheriting a "conformal geometry" (and an orientation) from the standard Euclidean geometry of the complex plane $\mathbb{C}$ via this natural mapping. A *conformal geometry* on a smooth surface is a "geome-

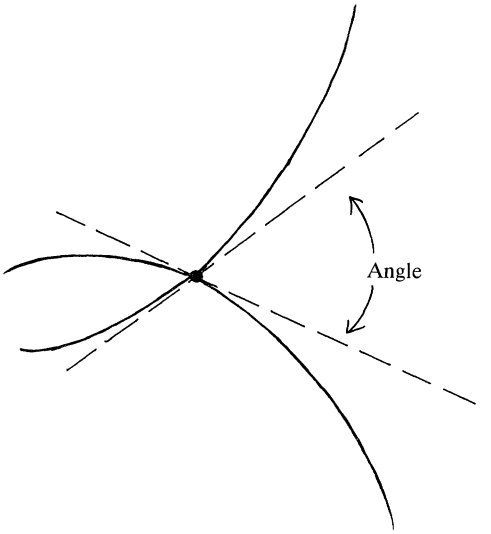


FIG. 9. In conformal geometry, there is no invariant notion of "length" of an arc, but there is a notion of angle.

try" where there is *no* intrinsic notion of "length of arc," but if we are given two, real, smooth pieces of arc traced on the surface which intersect at a point (as in FIGURE 9) it makes sense to talk about their *angle of intersection*.

To endow a smooth surface with an orientation and a (locally Euclidean) conformal geometry is the same as to consider it as a *Riemann surface*.[13] {Here is a sketch of the modern formulation of this notion, originating essentially with Herman Weyl: to give a smooth surface the structure of *Riemann surface* has to specify at each point of the surface a preferred class of complex-valued functions defined locally about the point, called *analytic functions*, where this specification is done in the following manner: Among the analytic functions about each point there is at least one, call it z, which identifies some neighborhood of the point with an open set U of the complex plane so that the specified class of *analytic functions* on that neighborhood is identified with the class of standard analytic functions on U. Moreover, the condition of being an 'analytic function' is a *local condition* in the sense that a function defined on an open subset of the surface is 'analytic' if and only if it is 'analytic' on each subset of any open cover of that open subset. Given, then, a Riemann surface we may say what it means for a complex-valued function on it to be 'analytic,' or 'meromorphic' (a ratio of 'analytic functions').

A function which plays the role required of the function denoted z above, which is defined locally about a point of the surface and, moreover, which vanishes at that point is sometimes called a *local uniformizer* at the point in question. The local isomorphism between an open subset of a Riemann surface and $U \subset \mathbb{C}$ given by a 'local uniformizer' z identifies the "conformal geometry" on that open subset with the Euclidean conformal geometry on U}.

With Weierstrass let us ask for (*doubly*) *periodic*[14] functions on the complex plane: functions which are periodic with respect to both T_{λ_1} and T_{λ_2}. These doubly periodic functions are precisely the meromorphic functions on the quotient space $\mathbb{C}/\Lambda$, meromorphic in the sense of its Riemann surface structure. As Weierstrass showed, the problem of constructing all such periodic functions can be elegantly settled by the construction of one special periodic function, the so-called $\wp$-function attached to $\Lambda, \wp_\Lambda(z)$ (call it $\wp$ for short)[15] which satisfies a nonlinear differential equation of the form

$$(\wp')^2 = 4 \cdot \wp^3 + A \cdot \wp + B,$$

where A and B are complex numbers, and $\wp'(z) = d\wp(z)/dz$. This turns out to be quite an adventitious mathematical construction! An unexpected number of problems are solved in a single stroke: First, the stated problem finds its solution in this construction, *for any doubly periodic (with respect to* Λ*) meromorphic function can be obtained as a rational function in* $\wp_\Lambda$ *and* $\wp'_\Lambda$. But it is also the case that any equation

$$Y^2 = 4 \cdot X^3 + A \cdot X + B \qquad (A, B \in \mathbb{C}) \qquad (*)$$

[13]See Felix Klein's classic expository account: On Riemann's Theory of Algebraic Functions and Their Integrals (A Supplement to the Usual Treatises). Dover.

[14](meromorphic)

[15]given by: $\wp_\Lambda(z) = 1/z^2$ plus the infinite sum of the terms $\{1/(z-\lambda)^2 - 1/\lambda^2\}$ where λ ranges through the nonzero elements of the lattice Λ.

[such that the cubic polynomial $4 \cdot X^3 + A \cdot X + B$ has no multiple roots] arises as the polynomial relating $\wp_\Lambda$ and $\wp'_\Lambda$ where $\wp_\Lambda$ is the Weierstrass $\wp$-function attached to *some* lattice Λ. The algebraic curve in the (X, Y)-plane defined by such an equation (or rather, the completion of this curve in the projective plane) is called an *elliptic curve*. There is a single point at ∞ on this curve; this point will be called the *origin* of the curve. It is an easy exercise in algebra to see that *any* smooth algebraic curve of degree 3 in the projective plane can be given, after any choice of an origin and an appropriate rational change of variables, a defining equation of the above form (called a *Weierstrass equation*). This means that if we are interested in the *algebraic* problem of finding all complex solutions of smooth homogeneous forms in three variables of degree 3 (or equivalently, of finding solutions of a "Weierstrass equation") we are led to the *analytic* construction of a specific lattice Λ in $\mathbb{C}$, of its associated $\wp$-function $\wp_\Lambda(z)$, and then the non-origin points on E are given by $(X, Y) = (\wp_\Lambda(z), \wp'_\Lambda(z))$ where z ranges through the points of $\mathbb{C}$ which are not in Λ. Let us call the mapping

$$\mathbb{C} - \Lambda \xrightarrow{\mathscr{U}} \textit{elliptic curve} - \textit{origin}$$
$$z \mapsto (X, Y) = (\wp(z), \wp'(z))$$

a (*Euclidean*) *uniformization* of our elliptic curve. One can, if one wishes, complete the picture to get a mapping

$$\mathbb{C} \xrightarrow{\mathscr{U}} \text{elliptic curve}$$

which identifies the complex points of our elliptic curve (including its origin) with the orbit space with respect to Λ, i.e., the Riemann surface[16] $\mathbb{C}/\Lambda$, the mapping being the covering mapping.

The word "uniformization" is meant to carry the full load of its nuances here.

We have *uniformly* parametrized the complex points of *any* elliptic curve by the points of $\mathbb{C}$ (more exactly, if you wish, by the orbits of these points under translation via the elements of a lattice Λ determined by the elliptic curve).

The uniformization $\mathscr{U}$ also uniformizes the conformal geometry of any particular elliptic curve, in the sense that $\mathscr{U}$ identifies the conformal geometry of the Riemann surface, locally, with Euclidean conformal geometry (its inverse, locally defined, gives *local uniformizers* in the sense described above)[17].

So far, our equation $(*)$ has complex coefficients A, B. Number theory will properly enter our picture when we consider elliptic curves $Y^2 = 4 \cdot X^3 + A \cdot X + B$ where the coefficients A, B are *algebraic numbers*, or more specifically, *rational numbers*. To focus on the latter case which will be our eventual particular interest, let us refer to an elliptic curve whose coefficients A, B lie in $\mathbb{Q}$ as simply

[16] It has mystified generations of students that (algebraic) *curves* can be viewed as (Riemann) *surfaces*. The clash in terminology indicates that, as its name implies, Algebraic Geometry is at the meeting-ground between algebra and geometry: one thinks, for example, of $\mathbb{C}$ as the affine line (i.e., a *curve*) if one is thinking algebraically and as the complex plane (i.e., a *surface*) if one is thinking geometrically.

[17] The uniformization $\mathscr{U}$ has the further "uniformity" property, usually incorporated in the technical definition of a covering mapping, that any small enough disc on the elliptic curve has the property that its pre-image under $\mathscr{U}$ consists in a disjoint union of components, each of which maps isomorphically onto the given disc.

an *arithmetic elliptic curve*. But before we deal with arithmetic elliptic curves we have some hyperbolic geometry to do.

(III). Periodicity on the (non-Euclidean) hyperbolic plane—the setting for the classical theory of modular functions.

Let us turn now to *hyperbolic geometry*, the (independent) discovery of Bolyai, Gauss, and Lobachevsky.

Hyperbolic geometry is a homogeneous geometry satisfying all the Euclidean axioms except for the fifth postulate, and possessing *many* lines through a given point, parallel to a given line; it now has a number of equivalent concretizations. The model particularly useful to us is the *upper half-plane model*.

Here the points of the geometry are the points $z = x + iy$ in the upper half of the complex plane **H**, i.e., x can be any real number and y any *positive* real. The lines are either vertical straight lines $\{a + iy\}$ for a fixed real number a and all positive reals y, or else they are semi-circles abutting on the real axis. The upper half-plane model inherits a Riemann surface structure, and hence also a *conformal geometry* by virtue of its being an open subset in $\mathbb{C}$.

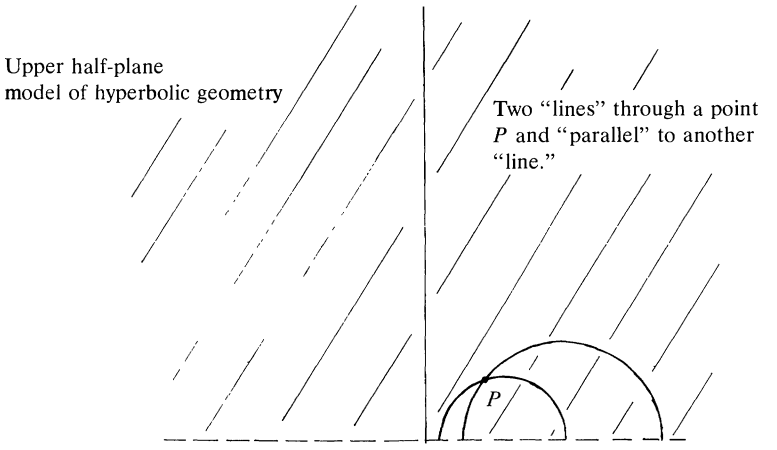


FIG. 10

The *translations* $T_b : z \mapsto z + b$ for any real number b are symmetries of hyperbolic geometry, but there are many more symmetries (in fact two other continuous parameters of them[18]), perhaps the most important single one being *inversion* with respect to the unit circle, $S : z \mapsto -1/z$.

[18]Consider matrices of real numbers of determinant equal to 1, i.e.,

$$\begin{pmatrix} a, b \\ c, d \end{pmatrix}$$

with $ad - bc = 1$. Then $z \mapsto az + b/cz + d$ is an orientation-preserving transformation of the upper half plane which is a symmetry of its hyperbolic geometry, and any orientation-preserving symmetry is given by such a matrix.

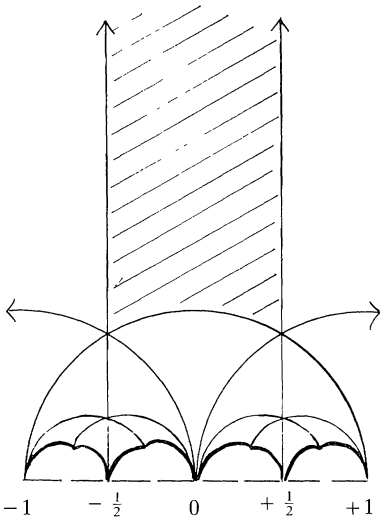


FIG. 11

FIGURE 11 is meant to illustrate the tiling of the hyperbolic plane that is gotten by systematically applying composites of iterates of the *unit translation*, $T_1 : z \mapsto z + 1$, and of the *inversion* S (and of their inverses), to the "basic tile," which is the shaded region in the figure. Let Γ be the group of symmetries of the hyperbolic plane gotten from such compositions of T_1 and S. It is a fact that Γ consists in *all* transformations of the form $z \mapsto az + b/cz + d$ where the coefficients a, b, c, d are all integers and $ad - bc = 1$. There are a number of striking differences between Γ acting on the hyperbolic plane and a lattice Λ, generated by translations T_{λ_1} and T_{λ_2}, acting on the complex plane. First, the two translations of the complex plane T_{λ_1} and T_{λ_2} commute with one another, which is not the case for *translation* and *inversion* of the hyperbolic plane, i.e., Γ is a more interesting, noncommutative, group. And second, there is a natural way of *identifying* Λ with

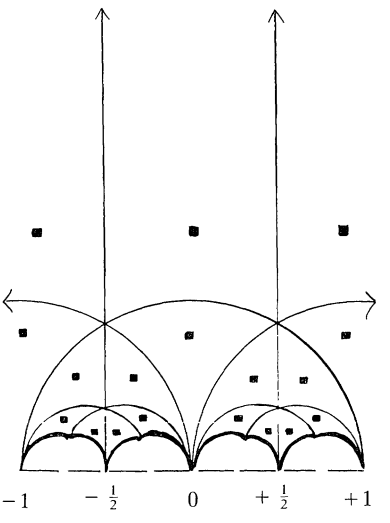


FIG. 12

an orbit in the complex plane (the orbit containing 0). There is no such identification for Γ[19]. In FIGURE 12, the "dots" (one in each tile) mark a single orbit with respect to Γ.

If E is an elliptic curve given with its conformal geometry, it has, of course, its *Euclidean* uniformization coming from Weierstrass's theory as described in II above. By a *hyperbolic uniformization of* E is meant a mapping

$$H - \{\textit{finite union of } \Gamma\textit{-orbits}\} \xrightarrow{\mathscr{V}} E - \{\textit{finite set of points}\},$$

which is a covering mapping of the domain onto the range[20], which preserves orientation, is a local isomorphism of conformal geometries[21], and which is periodic with respect to a subgroup of finite index in Γ. This last condition just means that there is a subgroup $\Gamma' \subset \Gamma$ of finite index such that if $z \mapsto \gamma(z)$ is a transformation in Γ', then

$$\mathscr{V}(\gamma(z)) = \mathscr{V}(z).$$

In other terms, $\mathscr{V}$ factors through the orbit space of the upper-half plane under the action of Γ'.[22]

For any such Γ', there are only a finite number of distinct elliptic curves admitting a hyperbolic uniformization periodic with respect to Γ'. Moreover, it is a consequence of a theorem of Bely that an elliptic curve admits a hyperbolic uniformization *if and only if* it has a Weierstrass equation with coefficients A, B which are algebraic numbers, i.e., $A, B \in \overline{\mathbb{Q}}$.

3. The conjecture of Shimura-Taniyama-Weil. A *hyperbolic uniformization* of E was defined to be a covering mapping periodic with respect to *any* subgroup of finite index in Γ. There is, however, a specific class of subgroups of finite index in Γ which plays a *dominant role* in arithmetic. To explain why this class should play the role it does (and there are interesting geometric reasons for this) would lead us far afield. But the definition of the class is simple enough: For a positive integer N, let $\Gamma(N)$ denote the group of matrices

$$\begin{pmatrix} a, b \\ c, d \end{pmatrix},$$

where a, b, c, d are integers, $ad - bc = 1$, and which are congruent to the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ modulo N.

The groups $\Gamma(N)$ act as groups of symmetries of the hyperbolic plane (the above matrix acting by the rule $z \mapsto az + b/cz + d$), these groups all being subgroups of finite index in Γ. Any subgroup of Γ containing one of the $\Gamma(N)$'s is called a *congruence subgroup*. The terminology here is quite straightforward: a *congruence subgroup* of Γ is simply one that is definable by congruence conditions on the coefficients of the matrices representing its elements. Not all subgroups of finite index in Γ are congruence subgroups. (It is even true that, at least according

[19]There are two orbits which are slightly more degenerate than the rest.

[20]As in footnote 17.

[21]Equivalently: is an analytic (unramified) covering mapping of Riemann surfaces.

[22]The uniformization $\mathscr{V}$ extends to an analytic mapping of the upper half-plane onto the complement of a finite set of points in E.

to one natural way of counting them, relatively few subgroups of finite index are congruence subgroups.) But this gives rise to the key.

DEFINITION. *Let E be an elliptic curve.* A hyperbolic uniformization (*of E*) of arithmetic type *is a hyperbolic uniformization of the elliptic curve E which is periodic with respect to a congruence subgroup* $\Gamma' \subset \Gamma$.

Although (by Weierstrass) *any* elliptic curve admits a Euclidean uniformization (and, in fact with respect to a lattice $\Lambda \subset \mathbb{C}$ unique up to complex scalar change), and (by Bely) an elliptic curve admits a hyperbolic uniformization if and only if it can be defined by a Weierstrass equation with coefficients A, B *which are algebraic numbers*, the Shimura-Taniyama-Weil conjecture asserts, further, that

Any arithmetic *elliptic curve* (*i.e., any elliptic curve whose defining equation can be taken with coefficients in* $\mathbb{Q}$) *admits a* hyperbolic uniformization of arithmetic type.[23, 24]

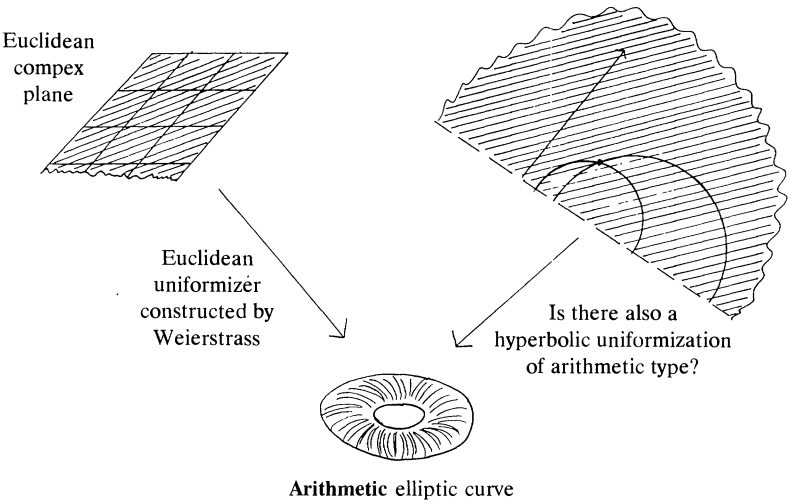


FIG. 13

[23]The formulation we have just given of the conjecture would make it seem "unfalsifiable." But in fact, there are more precise versions of the conjecture which predict, given an arithmetic elliptic curve E, exactly which $\Gamma(N)$ would be involved in a hyperbolic uniformization of arithmetic type for E—these precise versions are known (by the work of Hecke, Eichler, Shimura, Weil, Deligne, Carayol, Faltings, and others) to be equivalent to the one given here. A technical point relevant to this equivalence is treated briefly in an appendix to this expository article. There are also stronger conjectures by Langlands (concerning automorphic representations of reductive groups) and by Serre (concerning 2-dimensional representations of Galois groups over $\mathbb{Q}$) which imply the conjecture of Shimura-Taniyama-Weil.

[24]As Serre remarked, it might be illuminating to formulate a precise conjectural characterization of the class of elliptic curves (necessarily definable over $\overline{\mathbb{Q}}$) which admit hyperbolic uniformizations of arithmetic type. The conjecture of Shimura-Taniyama-Weil asserts, of course, that any elliptic curve definable over $\mathbb{Q}$ admits such a uniformization. Among the elliptic curves definable over quadratic number fields, for example, a *necessary* condition for them to have such a uniformization is that they be $\mathbb{C}$-isogenous to their conjugate (cf. Goro Shimura, Class fields over real quadratic fields and Hecke operators, 95 (1972) 130–190, where the case of real quadratic fields is analyzed and examples are given).

It is the confluence of *two* uniformizations, the Euclidean one, and the (conjectural) hyperbolic one of arithmetic type, that puts an exceedingly rich geometric structure on an arithmetic elliptic curve, and that carries deep implications for arithmetic questions.

4. Why does the Shimura-Taniyama-Weil conjecture imply Fermat's Last Theorem? This is not the place to go into the reams of evidence in favor of the conjecture, or of the overarching Langlands Program of which this conjecture is a minute part. But those readers who have borne patiently with me up to this point deserve some hints, however brief, of the connection between the conjecture which has been the focus of our talk and Fermat's Last Theorem.

Suppose, then, that we are given a *counterexample* to Fermat's Last Theorem, i.e., we are given a particular solution to a Fermat equation, $a^p + b^p = c^p$, where a, b, c are nonzero integers and the exponent p is an odd prime number. We might safely suppose that $p > 150{,}000$ since Fermat's last theorem has been established for odd exponents smaller than that bound. The subsequent argument, however, works for all primes $p \geqslant 5$, which is all that we assume at this point. We can always arrange the equation, by permuting (a, b, c) with appropriate changes of sign, if necessary, to get b to be even and a to be of the form $4k - 1$. Hellegouarch [**H**] and, more recently, Frey [**Fr** 1, 2] consider the arithmetic elliptic curve given by the cubic equation

$$y^2 = x(x - a^p)(x + b^p) \qquad (**)$$

which at first sight may not seem remarkable, but... (as is suggested in [**Fr** 2], formulated and set up in [**S**], and concluded in [**R**])

THEOREM. *The arithmetic elliptic curve* $(**)$ *does not admit a hyperbolic uniformization of arithmetic type.*

COROLLARY. *The conjecture of Shimura-Taniyama-Weil implies that the arithmetic elliptic curve* $(**)$ *does not exist, i.e., the conjecture of Shimura-Taniyama-Weil implies Fermat's Last Theorem.*

For a sketch of the proof of the above theorem, the reader might consult the Bourbaki report of Oesterlé [**O**]. The proof makes essential use of the classical and also the more modern arithmetic theory of modular forms. One supposes that $(**)$ does admit a hyperbolic uniformization of arithmetic type. Then, if ω is a regular differential 1-form on the elliptic curve, pulling ω back to the upper half-plane via the hyperbolic uniformization one gets a differential 1-form on the upper half-plane which after suitable normalization can be written as $f(z)\,dz$ where $f(z)$ is a cuspidal modular form f of weight 2 (the form f is modular for one of Hecke's groups and is an eigenform of the Hecke operators) with integral Fourier coefficients:

$$f(z) = 1 \cdot e^{2\pi iz} + \alpha_2 \cdot e^{4\pi iz} + \alpha_3 \cdot e^{6\pi iz} + \cdots \qquad (\alpha_j \in \mathbb{Z}).$$

Using an approach suggested by a conjecture of Serre, and using the specific form of the discriminant and the conductor of $(**)$[25], Ribet has shown that the

[25]The discriminant of the cubic equation $(**)$ is a *perfect pth power* times a power of 2 (it is precisely $(a^p b^p c^p / 2^8)^2$) and the conductor of $(**)$ is an even square-free number.

Fourier coefficients of $f, \alpha_1 = 1, \alpha_2, \alpha_3, \cdots$ are congruent modulo p to the Fourier coefficients of a modular form φ of weight two on Hecke's group $\Gamma_0(2)$. Ribet's argument is startling in its originality and makes use, among many other things, of the very mysterious "Drinfeld switch" which occurs in the description of the "bad fibers" of Shimura curves.

But by our good fortune there is only one modular form of weight two on $\Gamma_0(2)$, the Eisenstein series. That is, we know exactly what φ is, and therefore we know the Fourier coefficients of f, α_j modulo p.

Now, the theorem of Eichler-Shimura links the Fourier coefficients α_j modulo p to questions of $\mathbb{Q}$-rational p-torsion on the arithmetic elliptic curve $(**)$. From our explicit knowledge of α_j mod p, we deduce that there exists a rational point of order p on $(**)$[26].

But any arithmetic elliptic curve, which like $(**)$ has all four points of order 2 rational over $\mathbb{Q}$, cannot have any rational p-torsion for $p \geqslant 5$ (see [**M**]). Contradiction!

I hope, in these few minutes to have given some sense of the eclectic spirit of this conjecture, of how broadly it reaches out towards realms of mathematics that one might, at first, believe to be remote from Arithmetic, and yet how it gets to the heart of Arithmetic matters.

References for §4.

[**Fr** 1]. G. Frey, Rationale Punkte auf Fermatkurven and getwisten Modulkurven, *J. Crelle*, 331 (1982) 185–191.

[**Fr** 2]. _____, Links between stable elliptic curves and certain Diophantine equations, Ann. Univ. Saraviensis, Ser. Math. 1.

[**H**]. Y. Hellegouarch, Courbes elliptiques et équations de Fermat, Thèse, Bésancon, 1972.

[**M**]. B. Mazur, Modular curves and the Eisenstein ideal, *Publ. Math. I.H.E.S.*, 47 (1977) 33–186.

[**O**]. J. Oesterlé, Nouvelles approaches du "Théorème" de Fermat, Séminaire Bourbaki 87/88 *n* 694, *Astérisque*, 161–162 (1988) 165–186.

[**R**]. K. Ribet, On modular representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms, *Inv. Math.*, 100 (1990) 431–476.

[**S**]. J.-P. Serre, Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{Q}/Q)$, *Duke Math J.*, 54 (1987) 179–230.

Technical appendix: questions of fields of definition. These questions have not been dwelt upon, in the expository account above, because one of the points of my exposition was to emphasize the purely "geometric" nature of the Shimura-Taniyama-Weil conjecture. Nevertheless, the more standard expression of this conjecture (and, in fact, the form of the conjecture that has direct arithmetic consequences) is that given any elliptic curve E over $\mathbb{Q}$, there is a nonconstant mapping *defined over* $\mathbb{Q}$ from a modular curve $X_0(N)$ onto E. Here $X_0(N)$ is viewed as an algebraic curve over $\mathbb{Q}$ (given its "canonical model" structure over $\mathbb{Q}$, as initiated in [**Sh** 1]).

The point of this appendix is to make clear that we have not "cheated" in the statement we gave in the body of our expository account. That is, we will show that the "Shimura-Taniyama-Weil conjecture" as given in our expository text is equivalent to the statement in the preceding paragraph.

[26]Or at least on an arithmetic elliptic curve isogenous to $(**)$ via an isogeny of degree p rational over $\mathbb{Q}$.

Specifically, we shall sketch the proof that if an elliptic curve defined over $\mathbb{Q}$ admits a nonconstant mapping from $X(N)$ *defined over* $\mathbb{C}$, for some N, then it admits a nonconstant mapping from $X_0(N)$ *defined over* $\mathbb{Q}$ as well (but possibly for a *different* value of N). Now this fact is surely a "folk-theorem" either known to, or else easily derivable by, any of the experts. We record some brief indications of its proof here simply because it doesn't seem to be anywhere in the literature (but see [**Ri**] for very closely allied results).

Let K be a field (characteristic 0 is all we shall need) and consider the category $\mathscr{A}_K$ whose objects are abelian varieties over K, and whose morphisms are $\mathbb{Q}$-vector spaces obtained as follows: If A, B are abelian varieties over K, and $\mathrm{Hom}_K(A, B)$ is the abelian group of K-homomorphisms of abelian varieties, then the $\mathbb{Q}$-vector space of morphisms from A to B in $\mathscr{A}_K$ is $\mathrm{Hom}_K(A, B) \otimes \mathbb{Q}$. Thus, two abelian varieties are isomorphic in $\mathscr{A}_K$ if and only if they are K-isogenous. Let A be an abelian variety over K, and let M be a finite-dimensional $\mathbb{Q}$-vector space with continuous $\mathrm{Gal}(\overline{K}/K)$-action (i.e., the action factors through a finite quotient), where $\overline{K}$ is an algebraic closure of K. Then $A \otimes M$ is the unique object of $\mathscr{A}_K$ representing the contravariant functor $B \mapsto \{\mathrm{Hom}_{\overline{K}}(B, A) \otimes M\}^{\mathrm{Gal}(\overline{K}/K)}$ where the superscript $\mathrm{Gal}(\overline{K}/K)$ refers to the elements invariant under the (diagonal) action of $\mathrm{Gal}(\overline{K}/K)$ on $\mathrm{Hom}_{\overline{K}}(B, A) \otimes M$.

Let N be a positive integer, and $X_1(N)_{/\mathbb{Q}}$ the canonical model over $\mathbb{Q}$ of the modular curve attached to the classical group $\Gamma_1(N)$, as in [**Sh** 3]. Let $J_1(N)_{/\mathbb{Q}}$ be the jacobian of $X_1(N)_{/\mathbb{Q}}$, which we view as object of the category $\mathscr{A}_K$. Let E be an elliptic curve over $\mathbb{Q}$, and let $L/\mathbb{Q}$ be a field extension.

DEFINITION. *The elliptic curve E is called* modular ***over*** *L if there is an integer N and a nonconstant mapping $X(N)_{/L} \to E_{/L}$ (rational over L).*

PROPOSITION. *The elliptic curve E is modular over L if and only if E is modular over $\mathbb{Q}$ and more specifically, if and only if there is a nonconstant mapping $X_0(N)_{/\mathbb{Q}} \to E_{/\mathbb{Q}}$ (rational over $\mathbb{Q}$).*

Sketch of proof. We may assume that E does not have complex multiplication over $\mathbb{C}$, for if it did, then it would be modular over $\mathbb{Q}$ [**Sh** 4].

If E is modular over a given field, then it is also modular over any extension field, and if E is modular over $\mathbb{C}$ then it is modular over $\overline{\mathbb{Q}}$, so we need to prove that if E is modular over $\overline{\mathbb{Q}}$, then it is also modular over $\mathbb{Q}$. Recall that there is a nonconstant mapping $X_1(N^2) \to X(N)$ defined over $\overline{\mathbb{Q}}$, so (after a possible change of level N, and reduction from field of definition $\overline{\mathbb{Q}}$ to a suitable finite Galois extension $L/\mathbb{Q}$) we may assume that $E_{/\mathbb{Q}}$ is an elliptic curve without complex multiplications, admitting an L-rational nontrivial mapping $X_1(N) \to E$. Let $G = \mathrm{Gal}(L/\mathbb{Q})$.

By consideration of the Weil restriction from L to $\mathbb{Q}$ of this homomorphism of abelian varieties, and using the fact that $J_1(N)_{/\mathbb{Q}}$ is isogenous (over $\mathbb{Q}$) to a product of abelian varieties A_f for newforms f (Prop. 2.3 of [**Ri**]; for more about the varieties A_f see [**Sh 2, 3**]) we see that there is an irreducible G-module M and a newform f with Fourier coefficients in $\mathbb{C}$ such that A_f is isomorphic in the category $\mathscr{A}_{\mathbb{Q}}$ to $E \otimes M$ (i.e., these abelian varieties are isogenous over $\mathbb{Q}$). Let $F \subset \mathbb{C}$ be the coefficient field of the newform f (compare [**Ri**]). Then $[F : \mathbb{Q}] = \dim A_f = \dim_{\mathbb{Q}}(M)$. We also have that $F = \mathrm{End}_{\mathscr{A}_{\mathbb{Q}}}(A_f)$, by (Cor. 4.2 of [**Ri**]).

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

3Q; 1A

HENRY S. WARREN, JR.
IBM Thomas J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598

Here is the answer; what is the question?

$$f(y) = \cos^{-1}\sqrt{y} + \sqrt{y(1-y)}\,. \tag{1}$$

In this note we give *three* quite natural questions that have this same unlikely answer.

Question 1. Two bodies, of masses m and M, are attracted by gravity. Find the distance y between them at time t. Assume that the bodies are initially at rest and separated by one unit of distance, that $M \gg m$, and that $GM = 1$, where G is the universal gravitational constant.

Solution. Since $M \gg m$, we may assume that M does not accelerate. By Newton's second law ($F = ma$) and his law of gravity,

$$F = ma = my'' = -\frac{GmM}{y^2}.$$

Since $GM = 1$, this becomes

$$y'' = -\frac{1}{y^2}.$$

Hence $y'y'' = -y'/y^2$, or $(1/2)((y')^2)' = (1/y)'$, i.e., we have the "energy integral"

$$\frac{1}{2}(y')^2 - \frac{1}{y} = const = -1.$$

Integrating once more yields, after some calculation,

$$\sqrt{2}\,t = f(y),$$

where f is given by (1).

Question 2. A wheel of unit diameter rolls to the right on the x-axis, without slipping. Find the locus of a point P on the rim that is initially at $(0, 1)$.

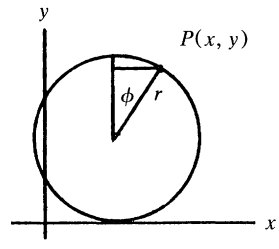


FIG. 1

Solution. The figure above shows the wheel after rotating an angle ϕ. It has moved along the x-axis an amount $r\phi$, so the center has moved this amount. From the figure,

$$x = r\phi + r \sin \phi, \tag{2a}$$

$$y = r + r \cos \phi. \tag{2b}$$

After some calculation, and taking $r = 1/2$, we have

$$x = f(y),$$

where f is given by (1).

Question 3. Find the perimeter P of the segment ABC of FIGURE 2a below, in terms of the distance y of the chord AC from point D on the opposite side of the circle. The circle has unit diameter.

Solution. Let O be the center of the circle, and draw auxiliary lines as shown in FIGURE 2b. Let ϕ denote the equal angles AOB and COB, and let r be the radius of the circle. From FIGURE 2b,

$$P = 2r\phi + 2r \sin \phi,$$

$$y = r + r \cos \phi.$$

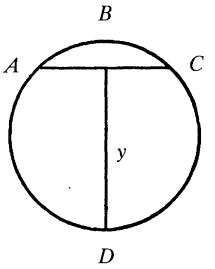


FIG. 2a

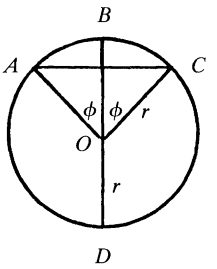


FIG. 2b

These equations are the same as equations (2) of question 2, with $P/2$ in place of x. Hence we have immediately

$$P/2 = f(y),$$

where f is given by (1). A similar problem is to find the area a of the region bounded by arc ABC and by lines AD and CD in the figure below.

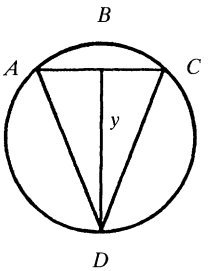


FIG. 3

The result is

$$2a = f(y),$$

where again f is given by (1).

An Occurrence of the Ballot Numbers in Operator Theory

LAWRENCE H. RIDDLE
Department of Mathematics, Agnes Scott College, Decatur, GA 30030

Let S denote the (unilateral) shift operator on the Hilbert space l_2 that maps the sequence $(x_1, x_2, x_3, \ldots)$ to the sequence $(0, x_1, x_2, x_3, \ldots)$. The adjoint operator S^* is a left shift, that is, $S^*(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots)$. Halmos has called the shift operator "the most important single operator, which plays a vital role in all parts of Hilbert space theory" [6]. It has many curious algebraic and analytic properties. It is an isometry of l_2 whose range is not the entire space. It has no eigenvalues while the eigenvalues of S^* comprise the open unit disk. In fact, the adjoint operator is universal in the sense that every operator whose spectrum is included in the interior of the unit disk is similar to a restriction of the adjoint operator to an invariant subspace. In this note we show how iterating commutators of the shift operator gives rise to the class of sequences known as the ballot numbers.

The ballot numbers, $[B_{i,j}]_{i,j=0}^{\infty}$, which play a role in a wide range of combinatorial arrangement problems [3], are defined by

$$\begin{cases} B_{0,0} = 1 & \\ B_{i,i+m} = 0 & m \geqslant 1 \\ B_{i,j} = B_{i-1,j-1} + B_{i,j+1} & i \geqslant 1, \end{cases} \tag{1}$$

where by convention $B_{i,-1} = 0$. One can interpret the integer $B_{i,j}$ as the number of ways to sequentially count votes given that you get i votes and your opponent gets $i - j$ votes and you are never behind in the tally. TABLE 1 gives the first several rows of the ballot numbers considered as an infinite lower triangular matrix.

TABLE 1: THE BALLOT NUMBERS

1								
1	1							
2	2	1						
5	5	3	1					
14	14	9	4	1				
42	42	28	14	5	1			
132	132	90	48	20	6	1		
429	429	297	165	75	27	7	1	
1430	1430	1001	572	275	110	35	8	1

The result is

$$2a = f(y),$$

where again f is given by (1).

An Occurrence of the Ballot Numbers in Operator Theory

LAWRENCE H. RIDDLE
Department of Mathematics, Agnes Scott College, Decatur, GA 30030

Let S denote the (unilateral) shift operator on the Hilbert space l_2 that maps the sequence $(x_1, x_2, x_3, \dots)$ to the sequence $(0, x_1, x_2, x_3, \dots)$. The adjoint operator S^* is a left shift, that is, $S^*(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots)$. Halmos has called the shift operator "the most important single operator, which plays a vital role in all parts of Hilbert space theory" [6]. It has many curious algebraic and analytic properties. It is an isometry of l_2 whose range is not the entire space. It has no eigenvalues while the eigenvalues of S^* comprise the open unit disk. In fact, the adjoint operator is universal in the sense that every operator whose spectrum is included in the interior of the unit disk is similar to a restriction of the adjoint operator to an invariant subspace. In this note we show how iterating commutators of the shift operator gives rise to the class of sequences known as the ballot numbers.

The ballot numbers, $[B_{i,j}]_{i,j=0}^{\infty}$, which play a role in a wide range of combinatorial arrangement problems [3], are defined by

$$\begin{cases} B_{0,0} = 1 & \\ B_{i,i+m} = 0 & m \geqslant 1 \\ B_{i,j} = B_{i-1,j-1} + B_{i,j+1} & i \geqslant 1, \end{cases} \tag{1}$$

where by convention $B_{i,-1} = 0$. One can interpret the integer $B_{i,j}$ as the number of ways to sequentially count votes given that you get i votes and your opponent gets $i - j$ votes and you are never behind in the tally. TABLE 1 gives the first several rows of the ballot numbers considered as an infinite lower triangular matrix.

TABLE 1: THE BALLOT NUMBERS

1								
1	1							
2	2	1						
5	5	3	1					
14	14	9	4	1				
42	42	28	14	5	1			
132	132	90	48	20	6	1		
429	429	297	165	75	27	7	1	
1430	1430	1001	572	275	110	35	8	1

TABLE 2

C_0:	1							
C_1:	1	−1						
C_2:	2	−3	1					
C_3:	5	−9	5	−1				
C_4:	14	−28	20	−7	1			
C_5:	42	−90	75	−35	9	−1		
C_6	132	−297	275	−154	54	−11	1	
C_7:	429	−1001	1001	−637	273	−77	13	−1

TABLE 3

$[C_0, S]$:	1							
$[C_1, S]$:	2	−1						
$[C_2, S]$:	5	−4	1					
$[C_3, S]$:	14	−14	6	−1				
$[C_4, S]$:	42	−48	27	−8	1			
$[C_5, S]$:	132	−165	110	−44	10	−1		
$[C_6, S]$:	429	−572	429	−208	65	−12	1	
$[C_7, S]$:	1430	−2002	1638	−910	350	−90	14	−1

diagonal element of C_i and let $[C_i, S]_j$ denote the jth off diagonal element of $[C_i, S]$. Then we have

PROPOSITION. *For all* $i, j \geqslant 0$,

(a) $$C_{i,j} = (-1)^j B_{i+j,2j}$$

(b) $$[C_i, S]_j = (-1)^j B_{i+j+1,2j+1}.$$

Proof. The proof for part (a) is by induction on i and makes use of the recurrence relations in equations (1) and (2). The case $i = 0, 1$ is true by direct verification. Now assume that (a) holds for an integer $i \geqslant 1$ and for all j. For $j = 0$ we have

$$C_{i+1,0} = C_{i,0} - C_{i,1} = B_{i,0} + B_{i+1,2} = B_{i+1,1} = B_{i+1,0}.$$

Now for $j \geqslant 1$,

$$\begin{aligned} C_{i+1,j} &= -C_{i,j-1} + 2C_{i,j} - C_{i,j+1} \\ &= -(-1)^{j-1} B_{i+j-1,2j-2} + 2(-1)^j B_{i+j,2j} - (-1)^{j+1} B_{i+j+1,2j+2} \\ &= (-1)^j \left[\left(B_{i+j-1,2j-2} + B_{i+j,2j} \right) + \left(B_{i+j,2j} + B_{i+j+1,2j+2} \right) \right] \\ &= (-1)^j \left[B_{i+j,2j-1} + B_{i+j+1,2j+1} \right] = (-1)^j B_{i+j+1,2j}, \end{aligned}$$

which completes the proof for part (a).

The proof for part (b) is almost identical once we realize that

$$[C_i, S]_j = -[C_{i-1}, S]_{j-1} + 2[C_{i-1}, S]_j - [C_{i-1}, S]_{j+1}.$$

Now use the same induction argument. ∎

In [**11**] it is shown that the ballot numbers are given by

$$B_{n,k} = \frac{k+1}{n+1}\binom{2n-k}{n}$$

for $0 \leqslant k \leqslant n$. From the proposition we thus obtain (after some simplification) that

$$C_{i,j} = (-1)^j \frac{2j+1}{2i+1}\binom{2i+1}{i-j} \qquad [C_i, S]_j = (-1)^j \frac{j+1}{i+1}\binom{2i+2}{i-j}.$$

In conclusion, we briefly mention two interpretations for the numbers in TABLE 2 and TABLE 3. The first concerns the enumeration of all possible pairings of $2n$ points on a circle by the number of crossings of the chords. Riordan [**9**] tabulates a table of coefficients for the enumerating generating functions for this problem, which, ignoring minus signs, is exactly TABLE 2. We refer the reader to [**9**] for more details. For the second interpretation, define a *path* to be a finite sequence of pairs $v_k = (a_k, b_k)$ of nonnegative integers such that (i) $v_0 = (0, 0)$; and (ii) if $v_k = (a_k, b_k)$, then $v_{k+1} = (1 + a_k, b_k)$ or $v_{k+1} = (a_k, 1 + b_k)$. A path $(v_0, v_1, \ldots, v_n)$ is of length n. The distance between two paths of length n is the absolute difference between the first coordinates of the last pair in each path. Finally, two paths of length n intersect if at least one set of corresponding pairs (other than $(0, 0)$) from each path are equal. Let $P_{i,j}$ denote the number of pairs of nonintersecting paths of length i and distance j. In [**10**] Shapiro exhibits a table of values for $P_{i,j}$, which he calls "a Catalan triangle," and studies their properties. Again ignoring minus signs, we have that TABLE 3 is exactly Shapiro's Catalan triangle.

Acknowledgement. The author would like to thank John D'Angelo for stimulating his initial interest in this problem, and the referees for suggestions that improved the content of the note.

REFERENCES

1. R. Alter, Some remarks and results on Catalan numbers, Proc. 2nd Louisiana Conf. on Combinatorics, *Graph Theory and Computing*, (1971) 109–132.
2. R. G. Eggleton and R. K. Guy, Catalan strikes again! How likely is a function to be convex, *Mathematics Magazine*, 61 (1988) 211–219.
3. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, John Wiley & Sons, Inc.
4. M. Gardner, Mathematical games, *Scientific American*, 234 (June 1976) 120–125.
5. H. W. Gould, Catalan Bibliography, Mathematics Department, West Virginia University.
6. P. Halmos, A Hilbert Space Problem Book, 2nd edition, Springer-Verlag.
7. F. T. Leighton and M. Newman, Positive definite matrices and Catalan numbers, *Proc. Amer. Math. Soc.*, 79 (1980) 177–181.
8. J. Riordan, Combinatorial Identities, John Wiley, New York, 1968, p. 62.
9. _____, The distribution of crossings of chords joining pairs of $2n$ points on a circle, *Math. of Computation*, 29 (1975) 215–222.
10. L. Shapiro, A Catalan triangle, *Discrete M.*, 14 (1976) 83–90.

11. _____, Positive definite matrices and Catalan numbers, revisited, *Proc. Amer. Math. Soc.*, 90 (1984) 486–496.
12. M. Shoaei and J. Sommerfeld, Catalan numbers in process synthesis, *AIChE Journal*, 32, 11 (1986) 1931–1933.

Derangements, Permanents, and Christmas Presents

STEPHEN G. PENRICE[1]
Department of Mathematics, Arizona State University, Tempe, AZ 85287

The derangement problem, also known as the hat-check problem, the matching problem, and "le probleme des recontres," was first solved by Montmort in 1713. There seem to be as many ways of describing the problem as there are names for it, and each description suggests some generalization, as several recent articles demonstrate (see **[2]**, **[3]**, and **[7]**). This note views the problem from a perspective that reaches back to my childhood. Because of the size of my family, we have traditionally held a drawing each year so that each child would be randomly assigned one brother or sister for whom he would buy a Christmas gift. Obviously, in order for such a drawing to be considered successful, no child can draw his own name. In mathematical language, we say that a derangement must occur. In any standard treatment of this problem, such as **[6]** or **[11]**, it is shown that as the number of children gets large, the probability that a successful drawing occurs tends to e^{-1}. The solution is a straightforward application of the inclusion-exclusion principle.

As with nearly everything else in life, however, the situation becomes more complicated as the children grow older. In our family, spouses of the children are included in the drawing, but this narrows the definition of a "successful drawing." In addition to having no one draw his own name, we want no one to draw his spouse's name, since presumably they exchange gifts anyway. Computing the probability of a successful drawing in this situation is difficult. In the case where all the children are married, we obtain bounds on this probability using two inequalities involving permanents of matrices, and these bounds are used to show that as the number of couples participating gets large the probability of a successful drawing goes to e^{-2}. We show this by proving a more general result.

Let k, n be integers, and S be a set of size kn. Let $T_1, T_2, \ldots, T_n$ be a partition of S into n sets, each of size k. Let $P_k(n)$ be the probability that a random permutation σ will have the property that for all x in S and for all i, $1 \leq i \leq n$, if x is in T_i, $\sigma(x)$ is not in T_i. The standard derangement problem is the case $k = 1$, and the "married" version of the "Christmas Gift Problem" is the case $k = 2$, with the couples involved constituting the partition. Our claim is that for k fixed, $\lim P_k(n) = e^{-k}$ as n goes to infinity. This result is implicit in example 7 on p. 177 of **[11]**, but it does not seem well known. The permanent inequalities mentioned above, known as Minc's conjecture and Van der Waerden's conjecture, allow us to give a simpler proof.

[1]Research supported in part by NSF grant DMS 87-01475 and ONR contract N00014-85K-0494 at the University of South Carolina.

First we define the permanent and show how it is related to our problem.

DEFINITION 1. Let $A = (a_{ij})$ be an $m \times m$ matrix. Define the permanent of A by

$$\operatorname{per}(A) = \sum_{\pi} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{m\pi(m)},$$

where Σ runs over all permutations π of $\{1, \ldots, m\}$.

While the permanent seems similar to the determinant, it is much harder to evaluate since it is not invariant under row and column operations. However, it is easy to verify that the permanent is multilinear, and this fact is all we need for our results.

Our plan now is to specify a (0, 1)-matrix $A_{k,n}$ such that $\operatorname{per}(A_{k,n})$ is exactly the number of permutations of the set S described above which map each set in the partition outside itself. Let $A_{k,n} = (a_{ij})$ where $1 \leq i, j \leq kn$ and where $a_{ij} = 0$ if and only if $q = q'$, where $i = qk + r$ and $j = q'k + r'$, $q, q', r, r' \in Z$, $0 < r$, $r' \leq k$, $0 \leq q$. $A_{k,n}$ is the $kn \times kn$ matrix with $k \times k$ blocks of 0's along the main diagonal and 1's everywhere else. For example, $A_{2,3}$ is

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

To see that $\operatorname{per}(A_{k,n})$ counts the permutations we seek, let $S = \{1, 2, \ldots, nk\}$ and let $T_1 = \{1, 2, \ldots, k\}$, $T_2 = \{k + 1, k + 2, \ldots, 2k\}, \ldots, T_n = \{(n - 1)k + 1, (n - 1)k + 2, \ldots, nk\}$. It is then clear that the non-zero terms of $\operatorname{per}(A_{k,n})$ correspond precisely to permutations which, for every i, map every x in T_i to an integer not in T_i.

As we noted earlier, permanents are difficult to compute in general. In fact there is a formula for $\operatorname{per}(A_{k,n})$:

$$\operatorname{per}(A_{k,n}) = \sum_{i=0}^{nk} (-1)^i R_i (nk - i)!,$$

where

$$R_i = \sum \prod_{j=1}^{n} \binom{k}{i_j} (k)_{i_j}$$

with the sum running over all $i_1 + i_2 + \cdots + i_n = i$, $0 \leq i_j \leq k$ and $(k)_i$ denoting the falling product $k(k - 1) \cdots (k - i + 1)$. We leave the derivation to the reader. (Hint: R_i counts the number of ways of selecting i zeroes from $A_{k,n}$ with no two in the same row or column. These R_i's are coefficients of a rook polynomial. For details on rook polynomials, consult chapter 7 of [**11**].) This formula does not seem to lend itself to straightforward analysis. However, Minc's Conjecture and Van der Waerden's Conjecture give bounds on $\operatorname{per}(A_{k,n})$ which are good enough to establish our result without analyzing this expression. We need two definitions to state these results.

DEFINITION 2. Let $A = (a_{ij})$ be an $m \times m$ matrix. The ith row sum of A is defined

$$r_i = \sum_{j=1}^{m} a_{ij}.$$

The jth column sum of A is defined

$$c_j = \sum_{i=1}^{m} a_{ij}.$$

DEFINITION 3. Let $A = (a_{ij})$ be an $m \times m$ matrix. Then A is doubly stochastic if

(i) $0 \le a_{ij} \le 1$ for $1 \le i, j \le m$

(ii) $r_i = 1$ for $1 \le i \le m$

(iii) $c_j = 1$ for $1 \le j \le m$.

Van der Waerden made his conjecture in [**13**], and the result was proved in [**4**] and [**5**]. Expositions of Egoritchev's proof can be found in [**8**] and [**14**].

THEOREM 4 (Van der Waerden's Conjecture). *Let A be a doubly stochastic $m \times m$ matrix. Then*

$$\operatorname{per}(A) \ge \frac{m!}{m^m},$$

with equality if and only if $a_{ij} = 1/m$ for all $1 \le i, j \le m$.

Minc's Conjecture was proposed in [**10**] and first proved in [**1**]. A simpler proof appears in [**12**].

THEOREM 5 (Minc's Conjecture). *Let A be an $m \times m$ $(0,1)$ matrix with nonzero row sums $r_1, \ldots, r_m$. Then*

$$\operatorname{per}(A) \le (r_1!)^{1/r_1} \cdots (r_m!)^{1/r_m}.$$

The next two lemmas are applications of these theorems. They are similar to results in [**9**] which give bounds on the number of perfect matchings in a regular bipartite graph.

LEMMA 6. *Let $A_{k,n}$ be as defined above. Then*

$$\operatorname{per}(A_{k,n}) \ge (kn)!(1 - 1/n)^{kn}.$$

Proof. Applying the multilinearity of permanents and Van der Waerden's Conjecture gives

$$\begin{aligned}
\operatorname{per}(A_{k,n}) &= (kn - k)^{kn} \operatorname{per}\left(\frac{1}{(kn - k)} A_{k,n}\right) \\
&\ge (kn - k)^{kn} \frac{(kn)!}{(kn)^{kn}} \\
&= (kn)!\left(1 - \frac{1}{n}\right)^{kn}.
\end{aligned}$$

LEMMA 7. $per(A_{k,n}) \leq ((kn-k)!^{1/kn-k})^{kn}$.

Proof. Apply Minc's Conjecture with $r_i = kn - k$ for all $1 \leq i \leq kn$.
We are now in a position to prove the result.

THEOREM 8. *Let k be fixed.* $\lim P_k(n) = e^{-k}$ *as n approaches infinity.*

Proof. By the above remarks $P_k(n) = \text{per}(A_{k,n})/(kn)!$. By lemmas 6 and 7

$$\left(1 - \frac{1}{n}\right)^{kn} \leq P_k(n) \leq \frac{\left(((kn-k)!)^{1/kn-k}\right)^{kn}}{(kn)!}.$$

It is easy to see that the lower bound approaches e^{-k}, and a straightforward calculation using Stirling's formula shows the upper bound is asymptotic to the lower bound.

REFERENCES

1. L. M. Bregman, Certain properties of non-negative matrices and their permanents, *Dokl. Akad. Nauk SSSR*, 211 (1973) 27–30. (Russian). (English translation: *Soviet Math. Dokl.*, 14 (1973) 945–949.)
2. B. D. Choi, Limiting distribution for the generalized matching problem, *Amer. Math. Monthly*, 94 (1987) 356–360.
3. K. David, Rencontres reencountered. *College Math. J.*, 19 (1988) 139–148.
4. G. P. Egoritchev, The solution of Van der Waerden's problem for permanents, *Adv. in Math.*, 42 (1981) 299–305.
5. D. I. Falikman, A proof of the Van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki*, 29 (1981) 931–938. (Russian). (English translation: *Math. Notes of the Academy of Science of the USSR*, 29 (1981) 475–479.)
6. W. Feller, An Introduction to Probability and its Applications, Vol. 1, 3rd. ed., Wiley, New York, 1968.
7. S. Fisk, The secretary's packet problem, *Math. Mag.*, 61 (1988) 103–105.
8. D. Knuth, A permanent inequality, *Amer. Math. Monthly*, 88 (1981) 731–740.
9. L. Lovasz and M. Plummer, Matching Theory, Akademiai Kiado, Budapest, 1986.
10. H. Minc, Upper bounds for permanents of (0,1)-matrices, *Bull. Amer. Math. Soc.*, 69 (1963) 789–791.
11. J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.
12. A. Schrijver, A short proof of Minc's Conjecture, *J. of Combin. Theory Ser. A*, 25 (1978) 80–83.
13. B. L. Van der Waerden, Aufgabe 45, *Jber. Deutsch. Math. Verein.*, 35 (1926) 117.
14. J. H. Van Lint, Notes on Egoritchev's proof of the Van der Waerden conjecture, *Linear Algebra and Its Appl.*, 39 (1981) 1–8.

Distance Functions and Topologies

FRED GALVIN*
Department of Mathematics, University of Kansas, Lawrence, KS 66045

S. D. SHORE
Department of Mathematics, University of New Hampshire, Durham, NH 03824

A *distance function* for a set X is a nonnegative, symmetric real-valued function d: $X \times X \to \mathbb{R}$ such that $d(p,q) = 0$ iff $p = q$. A *metric*, or *Euclidean* distance function, is a distance function that satisfies the triangle inequality: $d(p,q) \leq$

*Partially supported by NSF grant DMS-85-03732.

LEMMA 7. $per(A_{k,n}) \leq ((kn-k)!^{1/kn-k})^{kn}$.

Proof. Apply Minc's Conjecture with $r_i = kn - k$ for all $1 \leq i \leq kn$.
We are now in a position to prove the result.

THEOREM 8. *Let k be fixed.* $\lim P_k(n) = e^{-k}$ *as n approaches infinity.*

Proof. By the above remarks $P_k(n) = \text{per}(A_{k,n})/(kn)!$. By lemmas 6 and 7

$$\left(1 - \frac{1}{n}\right)^{kn} \leq P_k(n) \leq \frac{\left(((kn-k)!)^{1/kn-k}\right)^{kn}}{(kn)!}.$$

It is easy to see that the lower bound approaches e^{-k}, and a straightforward calculation using Stirling's formula shows the upper bound is asymptotic to the lower bound.

REFERENCES

1. L. M. Bregman, Certain properties of non-negative matrices and their permanents, *Dokl. Akad. Nauk SSSR*, 211 (1973) 27–30. (Russian). (English translation: *Soviet Math. Dokl.*, 14 (1973) 945–949.)
2. B. D. Choi, Limiting distribution for the generalized matching problem, *Amer. Math. Monthly*, 94 (1987) 356–360.
3. K. David, Rencontres reencountered. *College Math. J.*, 19 (1988) 139–148.
4. G. P. Egoritchev, The solution of Van der Waerden's problem for permanents, *Adv. in Math.*, 42 (1981) 299–305.
5. D. I. Falikman, A proof of the Van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki*, 29 (1981) 931–938. (Russian). (English translation: *Math. Notes of the Academy of Science of the USSR*, 29 (1981) 475–479.)
6. W. Feller, An Introduction to Probability and its Applications, Vol. 1, 3rd. ed., Wiley, New York, 1968.
7. S. Fisk, The secretary's packet problem, *Math. Mag.*, 61 (1988) 103–105.
8. D. Knuth, A permanent inequality, *Amer. Math. Monthly*, 88 (1981) 731–740.
9. L. Lovasz and M. Plummer, Matching Theory, Akademiai Kiado, Budapest, 1986.
10. H. Minc, Upper bounds for permanents of (0, 1)-matrices, *Bull. Amer. Math. Soc.*, 69 (1963) 789–791.
11. J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.
12. A. Schrijver, A short proof of Minc's Conjecture, *J. of Combin. Theory Ser. A*, 25 (1978) 80–83.
13. B. L. Van der Waerden, Aufgabe 45, *Jber. Deutsch. Math. Verein.*, 35 (1926) 117.
14. J. H. Van Lint, Notes on Egoritchev's proof of the Van der Waerden conjecture, *Linear Algebra and Its Appl.*, 39 (1981) 1–8.

Distance Functions and Topologies

FRED GALVIN*
Department of Mathematics, University of Kansas, Lawrence, KS 66045

S. D. SHORE
Department of Mathematics, University of New Hampshire, Durham, NH 03824

A *distance function* for a set X is a nonnegative, symmetric real-valued function d: $X \times X \to \mathbb{R}$ such that $d(p, q) = 0$ iff $p = q$. A *metric*, or *Euclidean* distance function, is a distance function that satisfies the triangle inequality: $d(p, q) \leq$

*Partially supported by NSF grant DMS-85-03732.

$d(p, x) + d(x, q)$ for any $x \in X$. Topologically, distance functions are interesting because they provide a convenient way of describing a topology and because their geometric nature often facilitates the understanding of this topological structure. It then becomes a question of *how* a distance function should generate a topology.

Intrinsically, a distance function describes the closeness of points in a set. Thus, for both a topology and a distance function, there is a concept of "neighborhood" of a point p. Consequently, it would seem that the topology determined by the distance function d is one for which the neighborhoods defined by d are the same as the neighborhoods defined by the topology. To this end, we let $B_d(p, \alpha)$ denote $\{x \in X | d(p, x) < \alpha\}$ and call this the open ball about p of radius α so that we have

DEFINITION 1. A *distance function* d *"determines" the topology* $\mathcal{T}$ provided that, for any point p, the set of open balls about p is a neighborhood base for p in the topological setting being described by $\mathcal{T}$; i.e, each open ball about p is a neighborhood of p and, when p is in any open set G in $\mathcal{T}$, then some open ball about p is also contained in G.

Historically, when a distance function determines a topology in this manner, it has been called a *semimetric*; see [**4**]. In this context we shall say that d determines the *semimetric topology* $\mathcal{T}$.

There is good reason to believe that this is the appropriate relationship between distance functions and topologies. In particular, note that, if d determines the semimetric topology $\mathcal{T}$, then a sequence x_n converges to a point p in the topological sense described by $\mathcal{T}$ if and only if the distance $d(x_n, p)$ converges to 0; i.e., the intrinsic convergence given by the distance function is the topological convergence. Conversely, if $\mathcal{T}$ is a first countable topology of X such that a sequence converges with respect to $\mathcal{T}$ iff it converges relative to a distance function d, then it follows that d is an appropriate distance function for determining the semimetric topology $\mathcal{T}$.

There is an alternative topological perspective for the way in which a distance function d determines a semimetric topology $\mathcal{T}$. For any distance function d, the *d-closure* of a subset A of X is $\{x \in X | d(x, A] = 0\}$, where $d(x, A] = \inf\{d(x, a) | a \in A\}$, the distance of x from A. One can easily show that d determines the semimetric topology $\mathcal{T}$ iff the d-closure of any subset A of X is identical to the topological closure of A with respect to $\mathcal{T}$. Finally, let us note that, if d is a distance function for X, then d determines some semimetric topology iff its d-closure is a topological closure operator (as defined in [**3**]).

Now that a relationship between a distance function and its topology has been established, we expect to develop certain properties (based implicitly on our Euclidean geometric experience), but there are some surprises in store.

EXAMPLE 1. There is a distance function d which determines a semimetric topology $\mathcal{T}$ and yet there are open balls generated by d that are not open.

Let $X = \{a_n | n = 1, 2, \ldots\} \cup \{0\}$ be the subset of $\mathbb{R}$ with its usual topology $\mathcal{T}$, where $a_n = 1/3^n$. Then the distance function d for X such that

$$d(x, y) = 1, \qquad \text{for } x \neq y \text{ and } xy \neq 0;$$
$$d(x, y) = |x - y|, \qquad \text{otherwise},$$

is a semimetric that determines $\mathcal{T}$. (It is a first countable space with the right

DEFINITION 2. A distance function d *determines the symmetric topology* $\mathcal{T}$ provided that a subset A of X is in $\mathcal{T}$ iff, when a point p is in A, then there is an open ball $B_d(p, \epsilon)$ about p contained in A.

It follows immediately that *any* distance function d will determine a unique symmetric topology $\mathcal{T}$ in which every open set is the union of open balls, even though the open balls about p need not be neighborhoods of p in the context of $\mathcal{T}$. In the case of Arhangel'skii's example, note that d determines a Hausdorff symmetric topology $\mathcal{T}$ such that a sequence x_n converges to p in $\mathcal{T}$ iff $d(x_n, p)$ converges to 0. Moreover, when the closure determined by d is a topological closure, then the symmetric topology is one for which the topological closure of a set coincides with its d-closure. Unfortunately, it seems that the desired convergence property has also been lost.

EXAMPLE 4. There is a distance function d which determines a symmetric topology $\mathcal{T}$ in which the topological convergence is not the same as the distance function convergence.

Let $X = A \cup B$, as defined in Example 2. There is a distance function d for X such that $d(b_n, a_m) = 1/3^n$, $d(b_n, 0) = 1$, and otherwise, $d(x, y) = |x - y|$. Note that the symmetric topology determined by d is one in which the sequence b_n converges to every point in A, including 0. However, $d(b_n, 0)$ does not converge to 0.

The failure to achieve identical convergence structures (for the topology and the distance function) in this example lies not so much in the connection between the distance function and the topology, but rather, in the pathological properties of the topological convergence itself. The sequence b_n is converging to almost anything!

Remark. If d determines a symmetric topology $\mathcal{T}$ in which convergent sequences have unique limits, then a sequence x_n converges to p iff $d(x_n, p)$ converges to 0.

Proof. Suppose that d is such a distance function. It is immediate that, if $d(x_n, p)$ converges to 0, then x_n converges to p in the symmetric topology $\mathcal{T}$ determined by d. Conversely, suppose that a_n converges to p, but $d(a_n, p)$ does not converge to 0. We may assume that there is a positive number ϵ such that $d(a_n, p) \geq \epsilon$ for every n. We claim that for each point q that is not in $A = \{a_n | n = 1, 2, \ldots\}$ there is an open ball about q that is disjoint from A. (This is obvious by our assumption, if $q = p$; otherwise, there is a subsequence of a_n that converges to q which is impossible if convergent sequences have unique limits.) Thus, A is closed and p is not in A, even though a_n converges to p. This contradiction establishes the converse.

Thus, it seems that we have finally arrived at the intrinsic connection (even for metrics) between a distance function and its topology.

REFERENCES

1. A. V. Arhangel'skii, Mappings and spaces, *Russian Math. Surveys*, 21 (1966) 115–162.
2. Fred Galvin and S. D. Shore, Completeness in semimetric spaces, *Pacific J. Math.*, 113 (1984) 67–75.
3. J. L. Kelley, General Topology, Springer-Verlag, New York, 1975.
4. W. A. Wilson, On semi-metric spaces, *Amer. J. Math.*, 53 (1931) 361–373.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A Characterization of Baire Spaces

DRAGAN JANKOVIĆ, MAXIMILLIAN GANSTER, AND IVAN REILLY
Department of Mathematics and Statistics, University of Auckland, Auckland, New Zealand

Completely metrizable topological spaces and locally compact Hausdorff spaces are two important classes of spaces that are different in many ways. However, each of these classes is contained in the class of Baire spaces, and this fact is one of the major motivations for the study of Baire spaces.

The term Baire space was introduced by Bourbaki to describe topological spaces in which every nonempty open subset is non-meager (that is, not the union of a countable family of nowhere dense subsets). The result that every complete metric space is a Baire space is known as the Baire Category Theorem. Early versions of it were proved independently by Osgood [4] in 1897 and Baire [1] in 1899. This theorem is one of the main ways that the notion of completeness is applied in analysis. Several of the important theorems of analysis, for example the Open Mapping Theorem, the Closed Graph Theorem and the Uniform Boundedness Theorem, require only that the space under consideration be a Baire space.

The purpose of this note is to present and discuss a characterization of Baire spaces which seems to be acknowledged in the literature only for the special case of complete metric spaces.

The fact that the set $\mathbb{Q}$ of rationals is not a G_δ-set (i.e., the countable intersection of open sets) in the real line $\mathbb{R}$ is well known and follows from the Baire Category Theorem for $\mathbb{R}$. Most students find that this application of the Baire Category Theorem is not trivial. Moreover, since $\mathbb{Q}$ is obviously meager in $\mathbb{R}$, it leads to the following natural question: What is the status of meager G_δ-sets in Baire topological spaces? The answer to this question is hardly surprising, but again is not obvious, namely, a topological space is a Baire space if and only if its meager G_δ-sets are nowhere dense. This observation is established in Kuratowski's book ([3], Theorem 2, p. 417) in the special case of complete metric spaces, but is missing in many standard textbooks on General Topology as well as in Haworth and McCoy's excellent monograph on Baire spaces [2], where some of its consequences are established. We believe the result deserves to be included among the standard characterizations of Baire spaces. The proof that we offer is very simple, and different from the proof of the special case in [3]. In order to present it let us recall a few concepts.

Let X be a topological space and let $A \subseteq X$. We denote the closure (or interior) of A in X by $\overline{A}$ (or $\operatorname{int} A$). Recall that a set $A \subseteq X$ is nowhere dense if $\operatorname{int} \overline{A} = \emptyset$. A subset of X is meager if it is a countable union of nowhere dense sets. As already mentioned we shall call a topological space X a Baire space if no nonempty open set in X is meager, i.e., meager subsets of X have empty interiors. Recall that a subset A of X is an F_σ-set if it can be represented as a countable union of closed subsets of X, that is A is a G_δ-set if and only if $X - A$ is an F_σ-set.

THEOREM. *A space X is a Baire space if and only if meager G_δ-sets in X are nowhere dense.*

Proof. For sufficiency, note that any meager nonempty open set would be a meager G_δ-set with nonempty interior, contradicting our condition. To show the necessity, let A be a meager G_δ-set in X. Then $X - A$ is an F_σ-set and hence $\bar{A} - A = \bar{A} \cap (X - A)$ is also an F_σ-set being the intersection of a closed set and an F_σ-set. Let $\bar{A} - A = \bigcup F_n$ be a representation of $\bar{A} - A$ as a countable union of closed sets. Since $\bar{A} - A$ has empty interior (this is true for any set A in X!), each F_n has empty interior and as a closed set with empty interior, F_n is nowhere dense. Therefore, $\bar{A} - A$ is meager and since the union of two meager sets is meager, $\bar{A} = A \cup (\bar{A} - A)$ is meager. Since X is a Baire space, $\operatorname{int} \bar{A} = \varnothing$ and hence A is nowhere dense. □

A topological space X is defined to be a T_1 space if whenever x and y are distinct points in X there is an open set containing x but not y. A space X is T_1 if and only if each singleton subset of X is closed. Recall that a point p in a space X is called isolated if the set $\{p\}$ is open in X.

COROLLARY 1 ([2], Proposition 1.29). *Let X be a Baire T_1 space with no isolated points, and let G be a countable G_δ-set in X. Then G is nowhere dense in X.*

Proof. Simply notice that countable sets in a T_1 space without isolated points are meager and apply the theorem. □

COROLLARY 2. *$\mathbb{Q}$ is not a G_δ-set in $\mathbb{R}$. Nonempty open sets in $\mathbb{R}$ are uncountable.*

COROLLARY 3 ([2], Proposition 1.30). *If X is a countable Baire T_1 space then the set of isolated points of X is dense in X.*

Proof. Observe that the set of non-isolated points of X is a meager G_δ-set and apply the theorem. □

The interested reader may check that Propositions 1.31 and 1.32 in [2] can also be easily established by use of the theorem above.

REFERENCES

1. R. Baire, Sur les functions de variables réelles, *Annali di Mat. Pura et Appl.*, 3 (1899) 65.
2. R. C. Haworth and R. A. McCoy, Baire spaces, *Dissertationes Math*, 141 (1977) 1–73.
3. K. Kuratowski, Topology, Vol. I, Academic Press, New York, 1966.
4. W. Osgood, Non-uniform convergence and the integration of series term by term, *Amer. J. Math*, 19 (1897) 155–190.

Which Latin Squares are Cayley Tables?

MAN-KEUNG SIU
Department of Mathematics, University of Hong Kong, Hong Kong

Visual display of a (finite) group as a Cayley table is usually introduced fairly early in an elementary course on abstract algebra. It is instructive in that several basic properties can be read off from the pattern of such tables. In particular, a

Cayley table is a latin square, i.e., the rows and columns are permutations of one another. However, the associative law is not easy to discern by the naked eye. A natural question is: *which latin squares are Cayley tables*?

We shall use $1, 2, \ldots, N$ to denote the entries in a latin square and $a(i, j)$ will denote the entry at the ith row and jth column. Below are two latin squares. Can you tell which one is a Cayley table (or not)?

1	2	3	4	5		1	2	3	4	5
2	5	4	1	3		2	4	1	5	3
3	1	2	5	4		3	1	5	2	4
4	3	5	2	1		4	5	2	3	1
5	4	1	3	2		5	3	4	1	2

Exhaustive checking involves N^3 pairs of triplets. M. Frolov [**4**] observed that a Cayley table possesses a necessary property, which turns out to be sufficient to guarantee that a latin square is a Cayley table. This is known as the "Quadrangle Criterion" [**3**, p. 18], viz. if $a(i, j) = a(i_1, j_1)$, $a(k, j) = a(k_1, j_1)$, $a(k, l) = a(k_1, l_1)$, then $a(i, l) = a(i_1, l_1)$. H. Brandt [**2**] showed that it suffices to assume $a(i, j)$ is the identity element. J. Aczél referred to the "Quadrangle Criterion" as Reidemeister Criterion and deduced from it one version of Brandt's criterion [**1**]. H. Zassenhaus [**6**, p. 4] and A. Hammel [**5**] phrased Brandt's criterion as a "Rectangular Rule."

In this note I wish to explain another means of checking which seems more convenient and which is based on the idea of the celebrated theorem of Cayley (1878) which says that every (finite) group is isomorphic to a group of permutations. Another reason for its pedagogical interest lies in this connection with Cayley's Theorem, which, basic though it is, is seldom *used* again in an elementary course after it has been stated and proved. Students see it, probably as an illustration in working with notions such as permutation, cancellation, isomorphism, but will not see it again (unless they study much later regular representation of a group in an advanced course). After all, it is no easy job to ferret out a subgroup of order N in the full symmetry group S_N with many more elements! Thus, it is comforting to be able to apply this idea in a situation which students can appreciate in an elementary course. (Towards the very end of [**5**], A. Hammel employs this same technique, but the explanation given is too brief and omits a main step of argument, especially in view of the pedagogical interest mentioned above.)

THEOREM. *A Latin square is a Cayley table if the composite of two rows is some row in the table.*

As an example, consider the two latin squares at the beginning. The one on the left is not a Cayley table, for the 2nd and 3rd rows yield a composite row (4 2 5 3 1) which is not a row of the square. The one on the right is seen to be a Cayley table after checking all composite rows. For instance, the 2nd and 4th rows yield a composite row (5 3 4 1 2), which is the 5th row of the square.

To prove the theorem, it remains to check if associative law holds in the set $G = \{1, \ldots, N\}$ with binary operation $*$ defined by the given latin square. We shall assume 1 is the identity element. Define φ from G into $A(G)$, which is the set of all bijections on G, by $\varphi(a)(x) = a * x$ for all x in G. The left cancellation law guarantees that $\varphi(a)$ is indeed in $A(G)$. Although G is a set for all we know, $A(G)$ is a group under composition. Associative law in G amounts to the relation

$\varphi(a) \circ \varphi(b) = \varphi(a * b)$ for all a, b in G. If that is true, then $\varphi(G)$ will be a subgroup of $A(G)$. Conversely, if $\varphi(G)$ is a subgroup of $A(G)$, then the associative law holds, for we claim that $\varphi(a) \circ \varphi(b)$, which is $\varphi(c)$ for some c in G, is indeed $\varphi(a * b)$. To see this, we shall show that c must be $a * b$. Towards this end, consider $\varphi(a) \circ \varphi(b)(1)$ and $\varphi(c)(1)$. The former is equal to $\varphi(a)(b * 1) = \varphi(a)(b) = a * b$, while the latter is equal to $c * 1 = c$. Hence, it reduces to check if $\varphi(G)$ is a subgroup of $A(G)$. Since $A(G)$ is a *finite* group we need only check if the composite of any two permutations in $\varphi(G)$ is still a permutation in $\varphi(G)$. This can be achieved by checking the rows of the latin square.

REFERENCES

1. J. Aczél, Conditions for a loop to be a group and for a groupoid to be a semigroup, *Amer. Math. Monthly*, 76 (1969) 547–549.
2. H. Brandt, Verallgemeinierung des Gruppenbegriffs, *Math. Ann.*, 96 (1927) 360–366.
3. J. Dénes, A. D. Keedwell, Latin Squares and Their Applications, *Akadémiai Kiado*, Budapest, 1974.
4. M. Frolov, Recherches sur les permutations carrées, *J. Math. Spéc.*, (3) 4 (1980) 8–11.
5. A. Hammel, Verifying the associative property for finite groups, *Math. Teacher*, 61 (1968) 136–139.
6. H. Zassenhaus, The Theory of Groups, Chelsea, New York, 1949.

An Overview of the Calculus Curriculum Reform Effort: Issues for Learning, Teaching, and Curriculum Development

JOAN FERRINI-MUNDY AND KAREN GEUTHER GRAHAM
Department of Mathematics, University of New Hampshire, Durham, NH 03824

1. Introduction to the calculus reform effort. Substantial discussion is underway in the mathematics education and mathematics communities concerning calculus pedagogy and curriculum [**6**, **9**, **28**] and there is considerable momentum for a national reform of the college calculus curriculum. Prominent mathematicians have elected to become involved in this process, and are aware of the immensity and importance of the enterprise. Young [**34**] comments: "Calculus is our most important course ... The future of our subject depends upon improving it." A brief summary of national data alone provides compelling documentation of the need for change. Annually, about 600,000 students enroll in calculus in four-year colleges and universities. About half of these students are enrolled in mainstream "engineering" calculus; of these 300,000, only 140,000 finish the year with a grade of D or higher. The average section size is 34 students; only 7% of calculus courses use computers; and exam questions require students to "solve, sketch, find, graph, evaluate, determine, and calculate" in a straightforward fashion. Surveys suggest

This article is based in part on work completed under a grant from the National Science Foundation (USE–8814057). Opinions, findings and conclusions expressed in this article are those of the authors and do not necessarily reflect the view of the National Science Foundation. Portions of this paper were presented at the 1989 Annual Meeting of the American Educational Research Association. The senior author is currently a Visiting Scientist in the Teacher Enhancement Program at the National Science Foundation; this paper was written in her role as a faculty member at the University of New Hampshire.

$\varphi(a) \circ \varphi(b) = \varphi(a * b)$ for all a, b in G. If that is true, then $\varphi(G)$ will be a subgroup of $A(G)$. Conversely, if $\varphi(G)$ is a subgroup of $A(G)$, then the associative law holds, for we claim that $\varphi(a) \circ \varphi(b)$, which is $\varphi(c)$ for some c in G, is indeed $\varphi(a * b)$. To see this, we shall show that c must be $a * b$. Towards this end, consider $\varphi(a) \circ \varphi(b)(1)$ and $\varphi(c)(1)$. The former is equal to $\varphi(a)(b * 1) = \varphi(a)(b) = a * b$, while the latter is equal to $c * 1 = c$. Hence, it reduces to check if $\varphi(G)$ is a subgroup of $A(G)$. Since $A(G)$ is a *finite* group we need only check if the composite of any two permutations in $\varphi(G)$ is still a permutation in $\varphi(G)$. This can be achieved by checking the rows of the latin square.

REFERENCES

1. J. Aczél, Conditions for a loop to be a group and for a groupoid to be a semigroup, *Amer. Math. Monthly*, 76 (1969) 547–549.
2. H. Brandt, Verallgemeinierung des Gruppenbegriffs, *Math. Ann.*, 96 (1927) 360–366.
3. J. Dénes, A. D. Keedwell, Latin Squares and Their Applications, *Akadémiai Kiado*, Budapest, 1974.
4. M. Frolov, Recherches sur les permutations carrées, *J. Math. Spéc.*, (3) 4 (1980) 8–11.
5. A. Hammel, Verifying the associative property for finite groups, *Math. Teacher*, 61 (1968) 136–139.
6. H. Zassenhaus, The Theory of Groups, Chelsea, New York, 1949.

An Overview of the Calculus Curriculum Reform Effort: Issues for Learning, Teaching, and Curriculum Development

JOAN FERRINI-MUNDY AND KAREN GEUTHER GRAHAM
Department of Mathematics, University of New Hampshire, Durham, NH 03824

1. Introduction to the calculus reform effort. Substantial discussion is underway in the mathematics education and mathematics communities concerning calculus pedagogy and curriculum [**6**, **9**, **28**] and there is considerable momentum for a national reform of the college calculus curriculum. Prominent mathematicians have elected to become involved in this process, and are aware of the immensity and importance of the enterprise. Young [**34**] comments: "Calculus is our most important course ... The future of our subject depends upon improving it." A brief summary of national data alone provides compelling documentation of the need for change. Annually, about 600,000 students enroll in calculus in four-year colleges and universities. About half of these students are enrolled in mainstream "engineering" calculus; of these 300,000, only 140,000 finish the year with a grade of D or higher. The average section size is 34 students; only 7% of calculus courses use computers; and exam questions require students to "solve, sketch, find, graph, evaluate, determine, and calculate" in a straightforward fashion. Surveys suggest

This article is based in part on work completed under a grant from the National Science Foundation (USE–8814057). Opinions, findings and conclusions expressed in this article are those of the authors and do not necessarily reflect the view of the National Science Foundation. Portions of this paper were presented at the 1989 Annual Meeting of the American Educational Research Association. The senior author is currently a Visiting Scientist in the Teacher Enhancement Program at the National Science Foundation; this paper was written in her role as a faculty member at the University of New Hampshire.

that students rarely are asked to deal with "higher level" problems, and that "pencil and paper algebraic manipulations are the order of the day" [**32**]. It is no small wonder that reform follows.

Davis observes [**8**, p. 36]: "When one sees large numbers of inexperienced teaching assistants presenting a 'notational' calculus to vast numbers of freshmen who may not be prepared to think about mathematics, it can seem that only the existence of non-conceptual 'manipulative' notational alternatives allows the whole enterprise to continue to operate ... These students show every evidence of having suffered from their pursuit of a meaningless, ritualistic manipulation of symbols."

The National Science Foundation has funded a number of projects that are intended to "explore opportunities for revitalizing instruction in calculus" [**21**, p. 2], and to support development of calculus curricula for the twenty-first century. Description of several of these efforts, and others not funded by NSF, is provided in the CRAFTY document, *Priming the Calculus Pump* [**6**]. Project emphases include: mathematical considerations, technological possibilities, and in some cases, beliefs or findings about student learning. Mathematical themes include optimization, estimation and approximation, differential equations, functions of several variables, dynamical systems theory, and numerical methods. Changes in the order of presentation of calculus topics, changes in emphasis within a given topic (generally toward a more conceptual emphasis), and increased attention to applications are appearing.

Several projects are based primarily on the availability of hardware and software that allow new possibilities for the teaching of calculus, including MAPLE, *Mathematica*, graphics packages, and symbolic/graphical calculators. Several projects display great enthusiasm for the potential of technology to ameliorate the calculus problems. Specifically, technology is proposed as a means of freeing students from algebraic manipulation, reducing the drudgery of calculation, supporting the learning of fundamental ideas, and allowing exploration of concepts.

We have identified a set of central issues worthy of careful consideration in this reform effort. We propose that those involved consider: potential orientations toward curriculum and curriculum development; the body of research available on student learning of calculus-related topics; the nature of the role of technology; and the role of teachers and instructors in the development and dissemination process. Each of these areas is discussed more fully below. Finally, we identify several research topics that are arising as part of this reform process.

2. Curriculum development and research on student learning. Much of the calculus reform discussion has centered on streamlining the traditional syllabus to create a "lean and lively" course [**10**]. Work available in mathematics education could be useful in this process [**17**]. For example, Romberg [**25**] characterizes curriculum as an "operational plan detailing what content is to be taught to students, how students are to acquire and use that content, and what teachers are to do in carrying out that curriculum." Traditional curriculum development is driven by content features, including organization of content into broad categories, segmenting of the categories, sequencing of topics, and specific lessons by topic. Romberg and Tufte [**26**] have argued for an approach that considers students' conceptions of mathematics, claiming that the prevailing conception is of mathematics as a static collection of concepts and skills to be mastered one by one. They

propose that "information about how individuals personally construct knowledge and store it in memory should be the basis of curriculum engineering."

Driver and Oldham's constructivist model [**12**] for curriculum development is very useful in considering the current reform. This model specifies four components that influence the design of curriculum: decisions on content, information about students' prior ideas, perspectives on the learning process, and teachers' practical knowledge. These components influence curriculum design, which consists of the design of learning strategies and materials. Implementation and evaluation follow, producing feedback which informs the original four components in an ongoing system of development.

Existing research on student learning in calculus has potential for strong impact on the development of calculus curriculum. A predominant theme in this area of research is that the learning of mathematics is a constructive process. This process occurs in the classroom context, with the learner attempting to make sense out of information by evaluating, connecting, and organizing it relative to prior experiences and existing knowledge structure. This line of research has challenged traditional assumptions about learning, such as the view of the student as a passive recipient of knowledge, or the notion that a misconception can be eradicated by the "perfect explanation" or by simply reorganizing the content. Research within this perspective suggests that knowledge about the variety and strength of student understandings must be taken into account when developing curriculum that seeks to facilitate learning and improve understanding.

Studies that have investigated student understanding of mathematical concepts use techniques such as the clinical interview, the teaching experiment, and the analysis of student errors. The use of qualitative techniques allows the researcher to generate a more viable model of the student's learning process than the more traditional quantitative methods [**33**]. A large portion of the research on student understanding has been conducted in the areas of early arithmetic, algebra, geometry, and physics [**1**, **2**, **7**, **18**, **24**, **29**].

In contrast, most studies in calculus have been large-scale, quantitative attempts to model how attitudinal and experiential variables affect overall performance [**9**, **13**, **14**, **15**, **20**]. Although several of these studies have provided valuable information, most lack a consistent perspective of student learning and have had minimal influence on understanding or performance.

A summary of the research in the area of student understanding of fundamental calculus concepts will be presented in the remainder of this section. The summary is divided into four areas: function, limits and continuity, the derivative, and the integral. In addition, several curriculum considerations suggested by this research will be discussed.

2.1 Function. Several researchers have investigated student understanding of the concept of function. This is one of the most fundamental concepts underlying the calculus. Despite its importance, students come to calculus with a primitive understanding of the concept and hold deeply rooted and firm misconceptions [**7**]. Students understand a function to be a formula; a graphical representation with no formula has no meaning for most entering calculus students [**11**]. In addition, their concept of function is static, in that they consider the function by thinking about one point at a time [**19**]. This is at odds with the traditional early expectation in

calculus that students understand secant lines "moving" on a curve to approach the tangent line, or are able to decide if "$f(x)$ is approaching some L as x approaches a."

Dreyfus and Eisenberg [**11**] suggest some additional characteristics of student understanding of function based on their interviews with algebra students. First, they observed a strong subservience to symbolism. Second, students have a strong attraction to linearity in that they will treat general functions as linear functions. Third, there is a notion that a relation is a function only when it can be represented by a single formula. Finally, students tend to view algebraic data and graphical data as being independent.

In our study involving clinical interviews with first-semester calculus students [**16**], we found that the students' understanding exhibited similar characteristics. Many of the students interviewed were not able to provide any type of general definition of function but readily gave examples of functions by writing formulas. There is little evidence that the students see functions as objects of study in mathematics; rather, when a function is given, in equation form, usually one is expected to do something to it, such as substitute in a value. This part of studying functions (plugging in values) seems to be firmly established and becomes their way of working with other calculus concepts such as limit.

2.2 Limits and continuity. Confrey [**4**] argues that students are likely to begin their study of calculus in one of three ways: with a discrete understanding of continuous ideas (time is a series of instants, motion a series of positions, etc.); with an independent transition to the idea of continuity; or with an algorithmic approach. Tall and Schwarzenberger [**30**] have investigated conflicts in the learning of limits and explored student responses to questions similar to those used by Confrey. They state that some of the difficulty arises from an information translation of very sophisticated ideas and use of the example of the definition of a sequence (s_n) of real numbers tending to a limit s: "Given any positive real number $\epsilon > 0$, there exists N (which may depend on ϵ) such that $|s_n - s| < \epsilon$ for all $n > N$." A typical informal translation is, "We can make s_n as close to s as we please by making n sufficiently large" [**30**, p. 44]. Tall et al. state that student difficulty may be inherent in the colloquial meaning of the terms used in this informal translation. For example, to many students the word "close" means near but not coincident. The student may reason from the informal translation that s_n can be close to, but not equal to, s. Other research in this area [**16**] suggests that students are quite successful at evaluating limits of the form $\lim f(x)$ as $x \to a$ where f is continuous at a. When given a simple limit problem in this format, with the appropriate graph displayed, most of the students in the study were able to solve the problem but showed very little geometric understanding. One of the students interviewed claimed that "the graph can't help me find an answer." Further probing revealed that the notion of "approaching" was not part of her understanding of limit. She saw limit problems as functions to be evaluated, and wasn't sure about all the "extra" notation (the arrow, the word "lim").

2.3 The derivative. Orton [**22**] conducted a large-scale qualitative study to investigate student understanding of differentiation and integration. Interviews were conducted with 110 students, 55 males and 55 females, between 16 and 22 years of age. The students were majoring in mathematics and had recently "sat

for" an examination that included elementary calculus. Orton found that the routine aspects of differentiation were well understood. All but four students were able to differentiate quadratic polynomials. A higher frequency of errors occurred when students were asked to differentiate a function of the form $y = a/x^2$. Most of the errors involved calculating the derivative of the quotient as the quotient of derivatives.

Answers on more conceptual tasks were far from satisfactory and the results suggest little "intuitive" understanding of the derivative, as well as some fundamental misconceptions.

Areas of student difficulty related to the tangent as the limit of a set of secants and to the ideas of rate of change of a straight line versus rate of change of a curve and rate of change at a point versus rate of change over an interval. According to Orton, the fundamental concept underlying rate of change is that of proportionality, which most of the students should have met in their study of graphs in elementary algebra. The results of this study suggest that the concept of proportionality may not be that elementary for a large number of students.

Orton suggests several implications for the curriculum and teaching based on these results. First, the basic concepts of calculus need to be revisited and reinforced at various times during the students' mathematics education. This in turn might mean stronger collaboration among secondary schools, colleges, and universities. Second, an initial approach to calculus concepts should be "informal" and involve numerical and graphical explorations. Orton states that real-life data should be employed, followed by a more algebraic development of the fundamental ideas. In addition, the role of technology in these explorations needs to be considered.

2.4 The Integral. Orton's study [**23**] also investigated student understanding of integration. Results suggest that students are able to apply, with some facility, the basic techniques of integration. However, further probing indicates that they possess fundamental misunderstandings about the underlying concepts. Orton found that the procedure of dissecting an area or volume, making use of the limit process, and the reasons why such a method works, were not part of most of the students' understanding of the integral. Activities need to be developed that explore the idea of limit in an intuitive way at all levels of the mathematics curriculum. Such explorations might involve exploring patterns, such as sets of polygons with an increasing number of edges, elementary number series, the Fibonacci sequence, and iterative solutions to equations [**23**]. Throughout these explorations the nature of the approximation process should be emphasized. Calculators and computers can serve as tools that aid such investigations. Orton comments, "the calculator does provide us with the opportunity of numerical approaches to calculus, and better understanding of the arithmetic may lead to better understanding of the algebraically equivalent procedures" [**23**, p. 11]. A curriculum is suggested that begins with an informal approach and becomes more formal with each new look.

In summary, research has found student understanding of central calculus concepts to be exceptionally primitive. Students demonstrate virtually no intuition about the concepts and processes of calculus. They diligently mimic examples and crank out homework problems that are predictably identical to the examples in the text. Misconceptions exist as a result of student attempts to adapt prior knowledge

having trouble. Order of presentation, emphasis of topics, and development of concepts are all more widely variable in the freshman calculus context than they are at earlier levels. Thus the problem of "national impact" is not trivial in any way.

In searching for a guide for thinking about this important issue for implementation, the experience of Cobb et al. [3] is especially helpful. The curriculum materials used in their research were developed in the context of the actual classroom, and were based on ongoing observations of children's mathematical activity in the classroom. This way of viewing the development of curriculum fully acknowledges the teacher's role in the implementation of curriculum, and could possibly be adapted as part of the calculus effort.

For any of the materials produced in these projects to become used in any widespread fashion, the dissemination process is indeed challenging. Whether the materials are organized with a novel mathematical perspective, as many are, or a new educational perspective, the dissemination should help convert people to the perspective. Given that materials may be used in a piecemeal fashion, when parts are appealing, the potential impact of this process will be to change people's way of thinking about the teaching and learning of calculus.

5. Research directions. A number of mathematics education research questions arise in conjunction with the calculus effort. Several questions relate to the scope and sequence of the mathematics content. Examples include: How does one decide which parts of the traditional curriculum can best be omitted? Is it more helpful to students to introduce the idea of limit before the idea of derivative? What are the effects of introducing substantial physical applications within the calculus course?

Many questions arise that relate directly to student learning and background matters. Examples of relatively broad questions include: Does lack of algebraic facility truly hinder calculus learning? How does the experience of secondary school calculus relate to the experience of college calculus? How does student "intuition" develop? Do physical examples help in the learning of concepts? How can underrepresented groups best be served in this revision process? How does procedural proficiency relate to conceptual understanding? Is a strong conceptual base adequate foundation for applying the ideas of calculus in new contexts within the sciences and engineering? There are also some examples of very specific learning questions which are rather intriguing, especially when raised by practicing mathematicians: Does a pointwise view of function interfere with the understanding of continuity on an interval? How natural and convincing is the traditional means of introducing the tangent line through the use of secant lines?

A third category of questions relates to pedagogical strategies and their interaction with mathematics content as well as student learning. Examples include: Will students be able to understand calculus concepts without extended manipulative practice? How does the use of writing enhance calculus learning? What are the relative benefits and disadvantages of large lectures? Can students learn calculus by working in pairs or groups? How can technology be helpful/detrimental with respect to student misconceptions? How do students use calculus textbooks, and what helpful features should be incorporated into textbooks?

Finally, there are questions related to teachers. How do teachers use textbooks and supplementary materials at the level of calculus? Will there be "teacher resistance" to widespread use of technology?

6. Conclusion. The calculus reform effort may prove to be an exciting phenomenon for those interested in mathematics education at this level. There clearly are opportunities for exciting work in curriculum, student learning, use of technology, and teacher development to inform this process in a productive fashion. In addition, possibilities for ongoing research that might both guide and result from this process seem rich and interesting.

REFERENCES

1. A. W. Bell, The learning process of aspects of mathematics, *Educational Studies in Mathematics*, 10 (1979) 361–387.
2. J. Clement, Algebra word problem solutions: Thought processes underlying a common misconception, *Journal for Research in Mathematics Education*, 13(1) (1982) 16–30.
3. P. Cobb, T. Wood and E. Yackel, Classrooms as learning environments for teachers and researchers, unpublished manuscript, Purdue University, 1989.
4. J. Confrey, Conceptual change, number concepts and the introduction to calculus, unpublished doctoral dissertation, Cornell University, Ithaca, N.Y., 1980.
5. J. Confrey and A. Lipton, Misconceptions research and the clinical interview. In S. K. Damarin & M. Shelton (eds.), Proceedings of the Seventh Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 44–49), Columbus, Ohio, 1985.
6. CUPM Subcommittee on Calculus Reform and the First Two Years, Priming the Calculus Pump: Innovations and Resources (MAA Notes Number 17), Mathematical Association of America, 1990.
7. R. B. Davis, Learning Mathematics: The Cognitive Science Approach to Mathematics Education, Norwood, N. J., Ablex Publishing Corportion, 1984.
8. R. B. Davis, Calculus at University High School. In R. G. Douglas (ed.), Toward a Lean and Lively Calculus (MAA Notes Number 6), Mathematical Association of America, 1986.
9. T. P. Dick, Covariance structural models for mathematics achievement and participation: An investigation of sex differences at the level of college calculus using factorial modeling (doctoral dissertation, University of New Hampshire, 1984).
10. R. G. Douglas (ed.), Toward a Lean and Lively Calculus (MAA Notes Number 6), Mathematical Association of America, 1986.
11. T. Dreyfus and T. Eisenberg, The function concept in college students: Linearity, smoothness, and periodicity, Unpublished manuscript, 1983.
12. R. Driver and V. Oldham, A constructivist approach to curriculum development in science, paper presented at the meeting of the British Educational Research Association, Sheffield, 1985.
13. O. P. Edge and S. H Friedberg, Factors affecting achievement in a first course in calculus. *Journal of Experimental Education*, 52(3) (1983) 136–140.
14. J. Ferrini-Mundy and R. Balomenos, Males and females in college calculus: Relationships among prior, concurrent, and subsequent academic experiences, Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1984.
15. K. J. Geuther, The role of error analysis, diagnostic grading procedures and student reflection in first semester calculus learning (doctoral dissertation, University of New Hampshire, 1986).
16. K. G. Graham and J. Ferrini-Mundy, An exploration of student understanding of central concepts in calculus, paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1989.
17. G. Howson, C. Keitel and J. Kilpatrick, Curriculum Development in Mathematics, Cambridge, England, Cambridge University Press, 1981.
18. C. Keiran, Concepts associated with the equality symbol, *Educational Studies in Mathematics*, 12 (1981) 317–326.
19. G. S. Monk, Students' understanding of functions in calculus courses, Unpublished paper, 1987.
20. J. Ferrini-Mundy, Spatial ability, mathematics achievement, and spatial training in male and female calculus students (doctoral dissertation, University of New Hampshire, 1980), Dissertation Abstracts International, 41, 4633A.
21. National Science Foundation Program Announcement and Guidelines. Undergraduate Curriculum Development in Mathematics: Calculus, Office of Undergraduate Science, Engineering, and Mathematics Education (USEME), Washington, DC, 1988.

22. A. Orton, Students' understanding of differentiation, *Educational Studies in Mathematics*, 15 (1983a) 235–250.
23. A. Orton, Students' understanding of integration, *Educational Studies in Mathematics*, 14 (1983b) 1–18.
24. F. Rief and J. I. Heller, Knowledge structure and problem solving in physics, *Educational Psychologist*, 17 (1981) 102–127.
25. T. A. Romberg, Curriculum, development and research, in M. F. Rosskopf (ed.), The Teaching of Secondary School Mathematics, Washington, D.C. National Council of Teachers in Mathematics, 1970.
26. T. A. Romberg and Fredric W. Tufte, Mathematics curriculum engineering: Some suggestions from cognitive science, in T. A. Romberg and D. M. Stewart (Eds.), The Monitoring of School Mathematics: Background Papers. Volume 2: Implications from Psychology; outcomes of instruction. Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison, 1987.
27. L. A. Steen, Forces for Change in the Mathematics Curriculum, text of an address given at a conference, The School Mathematics Curriculum: Raising National Expectations. UCLA, 1986.
28. L. A. Steen (ed.), Calculus for a New Century, background papers for a National Colloquium, October 28–29. National Research Council, 1987.
29. L. Steffe, Children's algorithms as schemes, *Educational Studies in Mathematics*, 14 (1983) 109–125.
30. D. O. Tall and R. L. Schwarzenberger (eds.), Conflicts in the learning of real numbers and limits, *Mathematics Teaching*, 83 (1978) 44–49.
31. W. P. Thompson, Experience, problem solving, and learning mathematics: considerations in developing mathematics curricula, in E. Silver (ed.), Teaching and Learning Mathematical Problem Solving (pp. 189–236), Hillsdale, N.J. Lawrence Erlbaum Associates, 1985.
32. T. W. Tucker, Calculus Tomorrow, in L. A. Steen (ed.), Calculus for a New Century (pp. 14–17). (MAA Notes Number 8), Mathematical Assocition of America, 1986.
33. E. von Glasersfeld, Learning as a constructive activity, in C. Janvier (Ed.), Problems of Representation in the Teaching and Learning of Mathematics (pp. 3–71). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1987.
34. G. S. Young, Present Problems and Future Prospects, in L. A. Steen (ed.), Calculus for a New Century (pp. 172–175). (MAA Notes Number 8). Mathematical Association of America, 1986.

An Early Introduction to Dynamics

MARYSIA T. WEISS
Mathematics Department, Hofstra University, Hempstead, NY 11550

1. Introduction. One of today's most exciting areas of mathematics is the study of dynamical systems. There are numerous unsolved problems and the field is extremely active. Not only mathematicians, but also ecologists, chemists, economists, and physicists have become involved in the field. The theory of dynamical systems is used in computer graphics, population models, and meteorology, to name a few. Many mathematicians feel that some knowledge of the subject is imperative. It is not difficult to accomplish this, since certain sophisticated topics can be learned rather easily at a very elementary level. In fact, they can be incorporated into the calculus sequence. Simple examples of dynamical systems may be shown to the student when introducing the composition of functions. Continuity and the Intermediate Value Theorem are the only tools necessary to prove a special case of a very beautiful theorem of Sarkovskii, thus making it accessible to a calculus student. The shift map, which is extremely important in the study of dynamics, can be used as an illustration of continuity on a more abstract

22. A. Orton, Students' understanding of differentiation, *Educational Studies in Mathematics*, 15 (1983a) 235–250.
23. A. Orton, Students' understanding of integration, *Educational Studies in Mathematics*, 14 (1983b) 1–18.
24. F. Rief and J. I. Heller, Knowledge structure and problem solving in physics, *Educational Psychologist*, 17 (1981) 102–127.
25. T. A. Romberg, Curriculum, development and research, in M. F. Rosskopf (ed.), The Teaching of Secondary School Mathematics, Washington, D.C. National Council of Teachers in Mathematics, 1970.
26. T. A. Romberg and Fredric W. Tufte, Mathematics curriculum engineering: Some suggestions from cognitive science, in T. A. Romberg and D. M. Stewart (Eds.), The Monitoring of School Mathematics: Background Papers. Volume 2: Implications from Psychology; outcomes of instruction. Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison, 1987.
27. L. A. Steen, Forces for Change in the Mathematics Curriculum, text of an address given at a conference, The School Mathematics Curriculum: Raising National Expectations. UCLA, 1986.
28. L. A. Steen (ed.), Calculus for a New Century, background papers for a National Colloquium, October 28–29. National Research Council, 1987.
29. L. Steffe, Children's algorithms as schemes, *Educational Studies in Mathematics*, 14 (1983) 109–125.
30. D. O. Tall and R. L. Schwarzenberger (eds.), Conflicts in the learning of real numbers and limits, *Mathematics Teaching*, 83 (1978) 44–49.
31. W. P. Thompson, Experience, problem solving, and learning mathematics: considerations in developing mathematics curricula, in E. Silver (ed.), Teaching and Learning Mathematical Problem Solving (pp. 189–236), Hillsdale, N.J. Lawrence Erlbaum Associates, 1985.
32. T. W. Tucker, Calculus Tomorrow, in L. A. Steen (ed.), Calculus for a New Century (pp. 14–17). (MAA Notes Number 8), Mathematical Asocition of America, 1986.
33. E. von Glasersfeld, Learning as a constructive activity, in C. Janvier (Ed.), Problems of Representation in the Teaching and Learning of Mathematics (pp. 3–71). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1987.
34. G. S. Young, Present Problems and Future Prospects, in L. A. Steen (ed.), Calculus for a New Century (pp. 172–175). (MAA Notes Number 8). Mathematical Association of America, 1986.

An Early Introduction to Dynamics

MARYSIA T. WEISS
Mathematics Department, Hofstra University, Hempstead, NY 11550

1. Introduction. One of today's most exciting areas of mathematics is the study of dynamical systems. There are numerous unsolved problems and the field is extremely active. Not only mathematicians, but also ecologists, chemists, economists, and physicists have become involved in the field. The theory of dynamical systems is used in computer graphics, population models, and meteorology, to name a few. Many mathematicians feel that some knowledge of the subject is imperative. It is not difficult to accomplish this, since certain sophisticated topics can be learned rather easily at a very elementary level. In fact, they can be incorporated into the calculus sequence. Simple examples of dynamical systems may be shown to the student when introducing the composition of functions. Continuity and the Intermediate Value Theorem are the only tools necessary to prove a special case of a very beautiful theorem of Sarkovskii, thus making it accessible to a calculus student. The shift map, which is extremely important in the study of dynamics, can be used as an illustration of continuity on a more abstract

level. I shall begin with a simple example of a dynamical system suitable for an elementary calculus student.

Suppose f is a function defined everywhere. Let f^n denote the composition of f with itself n times, that is, $f^2 = f \circ f$, $f^3 = f \circ (f \circ f)$, and so on.

The set of points $\{x_0, f(x_0), f^2(x_0), f^3(x_0), \ldots\}$ is called the *orbit* of x_0 under f. The point x_0 is called a *period-n* if $f^n(x_0) = x_0$ and n is the smallest such number.

An important area in the field of dynamical systems is precisely the study of orbits of points under different functions.

EXAMPLE. Let $f(x) = x^2$

Graphical analysis, as shown in FIGURE 1, may be used to "determine" the orbit of a point. Observe that $f(x) = x^2$ has 2 fixed points: 0 and 1. The points in the interval $(0, 1)$ all go to 0, while the points in $(1, \infty)$ all go to infinity. A student may try to look at the orbits of other elementary functions such as $2x + 4$, x^3, $\sin x$, e^x.

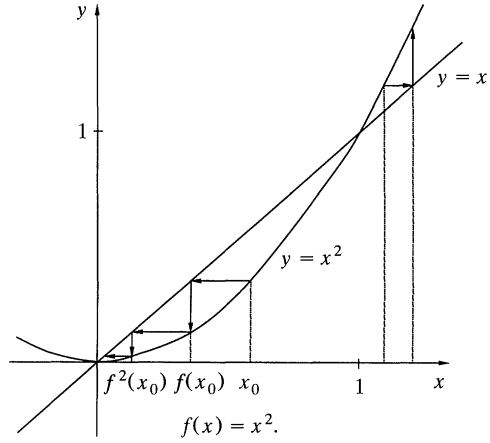


FIG. 1.

Now, consider a more general quadratic function: $f(x) = a(x + h)^2 + k$. It is interesting to see how the orbits of points are affected by changes in the parameters. One can ask a student to analyze graphically the orbits under f with specific values of a, h, and k. This turns out to be a nontrivial problem.

A well-known quadratic function, often used as a population model, is $F_\mu(x) = \mu x(1 - x)$ [**1**, pp. 31–38]. The dynamics of F_μ, a seemingly simple function, turn out to be quite variegated. It can be easily verified that for $\mu > 1$, F_μ has two fixed points, 0 and $p_\mu = (\mu - 1)/\mu$, where $0 < p_\mu < 1$, and that $F_\mu^n(x) \to -\infty$ as $n \to \infty$ for any $x < 0$ or $x > 1$. Thus, all the interesting dynamics of F_μ occur in the unit interval. If $1 < \mu < 3$, graphical analysis shows that $\lim_{n \to \infty} F_\mu^n(x) = p_\mu$ for all x in $(0, 1)$. The dynamics of F_μ become increasingly complicated as μ passes through 3. See FIGURES 2–5, where a solid line represents F_μ^3.

Once $\mu > 4$, certain points leave the unit interval after one iteration, since the maximum value of F_μ, $\mu > 4$, is larger than 1. See FIGURE 6.

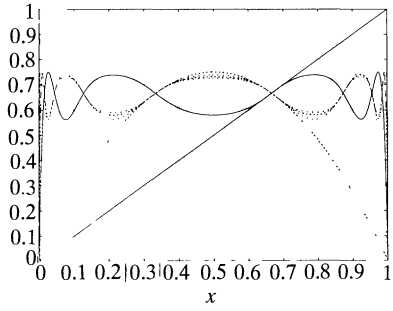


FIG. 2. Seven iterations of $3x(1 - x)$.

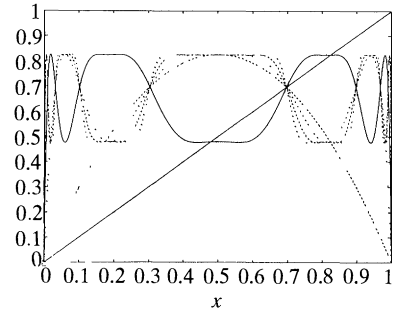


FIG. 3. Seven iterations of $3.3x(1 - x)$.

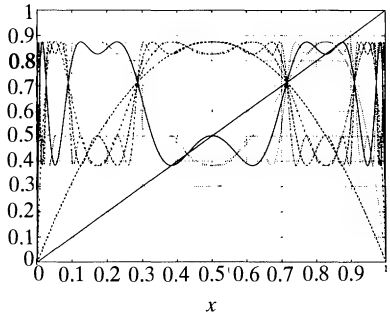


FIG. 4. Seven iterations of $3.5x(1 - x)$.

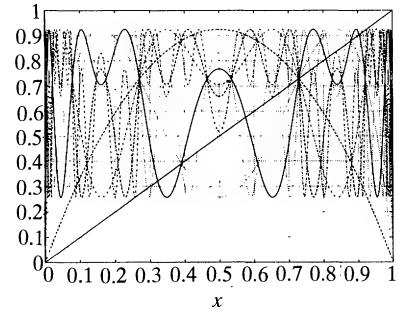
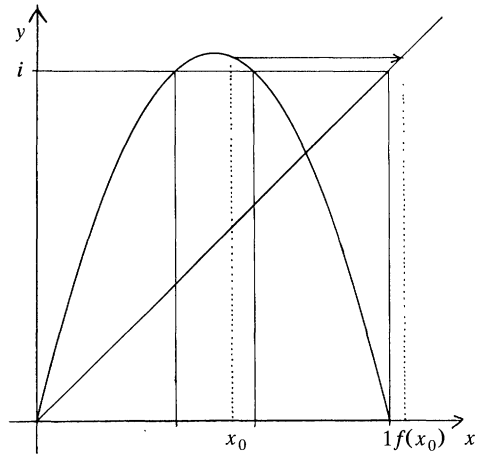


FIG. 5. Seven iterations of $3.7x(1 - x)$.



$F_\mu = \mu x(1 - x)$, $\mu > 4$.

FIG. 6. $F_\mu = \mu x(1 - x)$, $\mu > 4$.

Note that the only fixed points of σ are 0000... and 1111.... Also observe that 010101... and 101010... are the only period-2 points. What are the other periodic points of σ? Any period-n point has a repetitive "block" of n entries.

A simple example where the shift map is very useful involves a map of the circle. Let S^1 be the circle of radius one. We measure the angles in turns. Let f: $S^1 \to S^1$ be the map that doubles the angle from the x-axis. We can label the upper half of the circle by 0 and the lower half by 1.

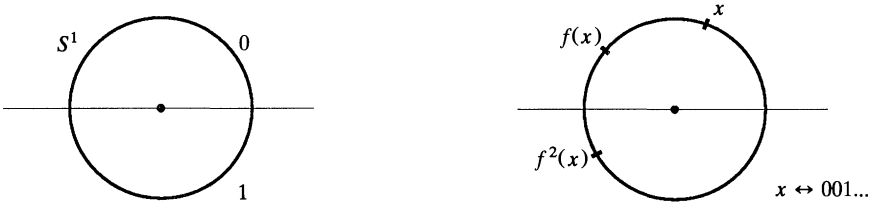


FIG. 7. f: $S^1 \to S^1$, an angle doubling map.

Now, to every point on S^1 we may associate a sequence in D analogous to the way such sequences are associated to numbers in [0, 1] via base 2 expansions. We define a function H: $S^1 \to D$ by $H(x) = s_0 s_1 s_2 \cdots$, where $s_i = 0$ if $f^i(x)$ is in the upper half of the circle, and $s_i = 1$ if $f^i(x)$ is in the lower half of the circle. Note that H is not a continuous function. Two points x_0 and x_1 may be arbitrarily close to each other, yet on the opposite sides of the half circle, with the result that $d(H(x_0), H(x_1)) \geqslant 1$. Nonetheless the map f can be viewed as the shift map. That is, the following diagram commutes.

$$\begin{array}{ccc} S^1 & \xrightarrow{f} & S^1 \\ {\scriptstyle H}\downarrow & & \downarrow{\scriptstyle H} \\ D & \xrightarrow[\sigma]{} & D \end{array}$$

It is much more difficult to show topological conjugacy of F_μ, where $\mu > 2 + \sqrt{5}$, and the shift map. The proof of this conjugacy is shown in [**1**, pp. 44–47].

3. Sarkovskii's theorem. In studying the dynamics of any function it is quite important to know its periodic points. There is a striking theorem of Sarkovskii dealing with periodic points of continuous real functions. A special case of Sarkovskii's Theorem, dealing with a period-3 point, was discovered independently by Li and Yorke [**3**] and later reproved by Straffin [**4**]. The general theorem, using Sarkovskii's ordering of positive integers, as well as the special case, are discussed in [**1**, pp. 60–68] and in [**2**]. Depending on the ability of the class, the special case of this theorem may be presented either in a strong intermediate or an advanced calculus course. The proof shown below is at a level suitable for an undergraduate. This early introduction to dynamics can spruce up the calculus sequence for faculty as well as for students.

THEOREM. *Suppose f is a continuous function whose domain and range is the set of real numbers. If f has a period-3 point, then it has period-n points for $n = 1, 2, \ldots$.*

Proof. Let a be a period-3 point of f. Let $b = f(a)$, $c = f(b)$. Without loss of generality we can assume that $a < b < c$. (If not, let $g(x) = -f(-x)$.)

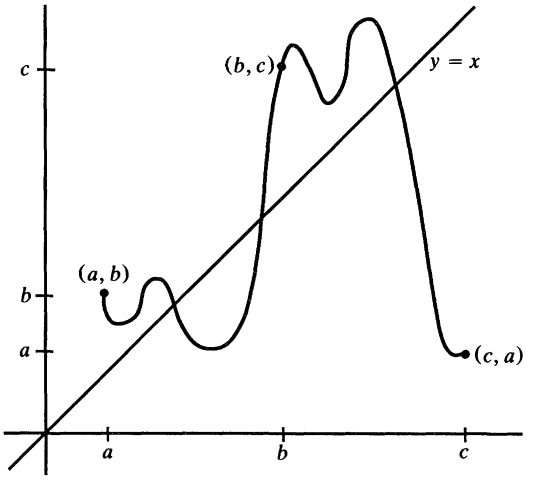


FIG. 8. A function with a period-3 point a.

It is geometrically clear that a continuous curve joining the points (b, c) and (c, a), $a < b < c$, must intersect the line $y = x$. This implies that there is at least one fixed point x of f in the interval $[b, c]$. More formally, the existence of the fixed point in $[b, c]$ follows from the Intermediate Value Theorem, since the image of the interval $[b, c]$ under f contains $[b, c]$. Thus we know that f has a period-1 point.

Now, $[a, b] \subset [a, c] \subset f^2([a, b])$. By a similar argument as above one concludes that there is at least one fixed point, x_1, of f^2 in $[a, b]$. Do we know, however, that there is at least one fixed point of f^2 which is not also a fixed point of f? By continuity and the facts that $f(b) = c$, and $c > b$, we can make a simple argument that there indeed is a genuine period-2 point in $[a, b]$ as follows. If there are fixed points of f in $[a, b]$ then there is a maximal fixed point $z_0 < b$. That is, if $x \in [a, b]$ is a fixed point, then $x \leqslant z_0 < b$, since $f(b) = c$ and f is continuous. Furthermore, by continuity of f we can choose $\varepsilon > 0$ small enough so that if $z_1 = z_0 + \varepsilon$ then $a < f(z_1) < b$. There are no fixed points of f in $[z_1, b]$. Let $f(z_1) = a_1$. Then $f([z_1, b]) \supset [a_1, c] \supset [b, c]$ and $f([b, c]) \supset [a, b]$; thus $f^2([z_1, b]) \supset [a, c] \supset [z_1, b]$. Therefore f^2 has a fixed point in $[z_1, b]$; that is, there exists $x_1 \in [a, b]$ such that $f^2(x_1) = x_1$ and $f(x_1) \neq x_1$.

Next let $H_0 = [a, b]$ and $H_1 = [b, c]$. To prove that there are period-n points for $n > 3$, we construct a sequence of intervals $\{I_i\}_{i=0}^{\infty}$ having the following two properties:

1. $H_1 = I_0 \supset I_1 \supset I_2 \supset I_3 \supset \ldots$
2. $f^i(I_i) = I_0$.

This construction is possible since the image of H_1 under f contains H_1, and thus there is a subinterval of H_1 whose image is exactly H_1. Proceeding inductively we get the desired nested sequence of intervals.

Now we shall construct another sequence of intervals, $\{J_i\}_{i=1}^{\infty}$. By the second property $f(I_1) = I_0$, so $f^2(I_1) = f(I_0) = f(H_1) \supset H$. Since $f^2(I_1) \supset H_0$, there is a subset of I_1, say J_1 such that $f^2(J_1) = H_0$, while $f(J_1) \subset H_1$. Similarly, there exists

$J_2 \subset I_2$, such that $f^3(J_2) = H_0$ while $f^2(J_2) \subset H_1$ and $f(J_2) \subset H_1$. We form a sequence of intervals $\{J_i\}_{i=1}^{\infty}$ such that:

1. $J_i \subset I_i$ for all $i > 0$
2. $f^i(J_{i-1}) = H_0$ while $f^j(J_{i-1}) \subset H_1$ for all $j < i$, $i > 1$.

Observe that $f^n(J_{n-2}) = f(H_0) \supset H_1 \supset J_{n-2}$. Therefore, f^n has a fixed point, p, in J_{n-2}; that is, $f^n(p) = p$. It is indeed a period-n point of f, since p, $f(p), f^2(p), \ldots f^{n-2}(p) \in H_1$, $f^{n-1}(p) \in H_0$ and $f^n(p) = p \in H_1$ again. If $f^{n-1}(p)$ is in the interior of H_0, then the proof is complete. If $f^{n-1}(p)$ is on the boundary, $f^{n-1}(p) = a$ or b, which implies that $p = f^n(p) = b$ or c. This, in turn, implies that $f(p) \notin H_1$ or $f^2(p) \notin H_1$ which indicates that $n = 2$ or 3, contradicting $n > 3$. □

It should be noted that the theorem has a very weak hypothesis while the conclusion is rather amazing. The existence of a period-3 point implies the existence of all other periods. The general theorem of Sarkovskii is even more powerful. This theorem involves a special ordering of the natural numbers, defined as follows:

$$3 \triangleright 5 \triangleright 7 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \cdots$$
$$\triangleright 2^3 \cdot 3 \triangleright 2^3 \cdot 5 \triangleright \cdots \triangleright 2^3 \triangleright 2^2 \triangleright 2 \triangleright 1.$$

SARKOVSKII'S THEOREM. *Suppose* $f\colon \mathbb{R} \to \mathbb{R}$ *is continuous and has a period-k point. If* $k \triangleright r$, *then f also has a period-r point.*

The function $F_\mu(x) = \mu x(1 - x)$ provides an example where both Sarkovskii's Theorem and a shift map is used. It can be shown by elementary computations that $F_{3.839}$ has a period-3 point. By Sarkovskii's Theorem $F_{3.839}$ must have periodic points of all other periods. Computations, however, do not reveal any other periodic points. A special type of a shift map is used to locate the other periodic points (see [**1,** §1.13]).

I would like to thank Professor David Knee for his helpful comments on this paper. I would also like to express my appreciation to the referees for several suggestions and to my student, Robert Ryan, for providing some of the graphs.

REFERENCES

1. Robert L. Devaney, An Introduction To Chaotic Dynamical Systems, second edition, Addison-Wesley, Redwood City, California, 1989.
2. W. de Melo and S. J. van Strien, Lectures on One-Dimensional Dynamical Systems, to appear.
3. T. Y. Li and J. A. Yorke, Period three implies chaos, this MONTHLY, 82 (1975) 986–988.
4. Philip D. Straffin, Jr. Periodic points of continuous functions, *Math. Mag.*, 51 (1978) 99–105.

A Simple Interpretation of the Complex Contour Integral

ALAN GLUCHOFF
Department of Mathematical Sciences, Villanova University, Villanova, PA 19085

Is there a simple interpretation for the familiar complex contour integral $\int_C f(z)dz$, one as accessible as those for the integrals of real-valued functions? My experience is that students are mystified on first exposure to this concept, and

$J_2 \subset I_2$, such that $f^3(J_2) = H_0$ while $f^2(J_2) \subset H_1$ and $f(J_2) \subset H_1$. We form a sequence of intervals $\{J_i\}_{i=1}^{\infty}$ such that:

1. $J_i \subset I_i$ for all $i > 0$
2. $f^i(J_{i-1}) = H_0$ while $f^j(J_{i-1}) \subset H_1$ for all $j < i$, $i > 1$.

Observe that $f^n(J_{n-2}) = f(H_0) \supset H_1 \supset J_{n-2}$. Therefore, f^n has a fixed point, p, in J_{n-2}; that is, $f^n(p) = p$. It is indeed a period-n point of f, since p, $f(p), f^2(p), \ldots f^{n-2}(p) \in H_1$, $f^{n-1}(p) \in H_0$ and $f^n(p) = p \in H_1$ again. If $f^{n-1}(p)$ is in the interior of H_0, then the proof is complete. If $f^{n-1}(p)$ is on the boundary, $f^{n-1}(p) = a$ or b, which implies that $p = f^n(p) = b$ or c. This, in turn, implies that $f(p) \notin H_1$ or $f^2(p) \notin H_1$ which indicates that $n = 2$ or 3, contradicting $n > 3$. □

It should be noted that the theorem has a very weak hypothesis while the conclusion is rather amazing. The existence of a period-3 point implies the existence of all other periods. The general theorem of Sarkovskii is even more powerful. This theorem involves a special ordering of the natural numbers, defined as follows:

$$3 \triangleright 5 \triangleright 7 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \cdots$$
$$\triangleright 2^3 \cdot 3 \triangleright 2^3 \cdot 5 \triangleright \cdots \triangleright 2^3 \triangleright 2^2 \triangleright 2 \triangleright 1.$$

SARKOVSKII'S THEOREM. *Suppose* $f\colon \mathbb{R} \to \mathbb{R}$ *is continuous and has a period-k point. If* $k \triangleright r$, *then f also has a period-r point.*

The function $F_\mu(x) = \mu x(1 - x)$ provides an example where both Sarkovskii's Theorem and a shift map is used. It can be shown by elementary computations that $F_{3.839}$ has a period-3 point. By Sarkovskii's Theorem $F_{3.839}$ must have periodic points of all other periods. Computations, however, do not reveal any other periodic points. A special type of a shift map is used to locate the other periodic points (see [**1,** §1.13]).

I would like to thank Professor David Knee for his helpful comments on this paper. I would also like to express my appreciation to the referees for several suggestions and to my student, Robert Ryan, for providing some of the graphs.

REFERENCES

1. Robert L. Devaney, An Introduction To Chaotic Dynamical Systems, second edition, Addison-Wesley, Redwood City, California, 1989.
2. W. de Melo and S. J. van Strien, Lectures on One-Dimensional Dynamical Systems, to appear.
3. T. Y. Li and J. A. Yorke, Period three implies chaos, this MONTHLY, 82 (1975) 986–988.
4. Philip D. Straffin, Jr. Periodic points of continuous functions, *Math. Mag.*, 51 (1978) 99–105.

A Simple Interpretation of the Complex Contour Integral

ALAN GLUCHOFF
Department of Mathematical Sciences, Villanova University, Villanova, PA 19085

Is there a simple interpretation for the familiar complex contour integral $\int_C f(z)dz$, one as accessible as those for the integrals of real-valued functions? My experience is that students are mystified on first exposure to this concept, and

working examples by the formula $\int_a^b f(z(t))z'(t)\,dt$ can be a baffling experience; what sense is a beginning student to make of the results

$$\int_{|z|=1} \operatorname{Re} z\,dz = \pi i \quad \text{or} \quad \int_1^i \frac{i}{z}\,dz = \frac{-\pi}{2}?$$

I would like to point out a way of looking at $\int_C f(z)\,dz$ which relies only on the geometry and algebra of complex numbers and can be introduced early in the discussion of integration. Using this idea, several examples will be worked out to show how values for some elementary integrals can be found.

Suppose a function f is defined on O, an open subset of $\mathbb{C}$ and C is a contour from A to B inside O. We will assume that $\int_C f(z)\,dz$ has been given a rough definition as $\lim \sum_{k=1}^n f(z_k^*)(z_k - z_{k-1})$, where $\{z_k\}_{k=0}^n$ are subdivision points of C, with $z_0 = A$ and $z_n = B$; z_k^* is a point on C between z_{k-1} and z_k, but will not worry about the details of convergence. Now if $L(C)$ is the length of C and $|z_k - z_{k-1}|$ is the distance from z_{k-1} to z_k in the complex plane, suppose $\{z_k\}_{k=0}^n$ are chosen so that $|z_k - z_{k-1}| = L(C)/n$, and that $z_k^* = z_{k-1}$; then we have

$$\begin{aligned}\sum_{k=1}^n f(z_k^*)(z_k - z_{k-1}) &= \sum_{k=1}^n f(z_{k-1})\frac{(z_k - z_{k-1})}{|z_k - z_{k-1}|}\cdot |z_k - z_{k-1}| \\ &= \frac{L(C)}{n}\sum_{k=1}^n f(z_{k-1})\frac{(z_k - z_{k-1})}{|z_k - z_{k-1}|},\end{aligned}$$

thus

$$\frac{1}{L(C)}\int_C f(z)\,dz = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f(z_{k-1})\frac{(z_k - z_{k-1})}{|z_k - z_{k-1}|}.$$

Now as $n \to \infty, (z_k - z_{k-1})/|z_k - z_{k-1}|$ is nearly equal to the "unit tangent vector" to C at z_{k-1}, that is, the complex number of modulus one whose vector representation is tangent to C at z_{k-1}. So, $(1/L(C))\int_C f(z)\,dz$ can be approximated by the average value of the product of f at z_k with the unit tangent vector at z_k, for $\{z_k\}$ a subdivision of C. Passing to the limit, one obtains $(1/L(C))\int_C f(z)\,dz$ = average for z on C of [$f(z)\cdot$ unit tangent at z], and if we write $T(z)$ = unit tangent to C at z, we have the formula

$$\frac{1}{L(C)}\int_C f(z)\,dz = \operatorname*{av}_{z\in C}[f(z)T(z)].$$

This is the proposed interpretation; we make a few comments before passing on to examples.

First, the average appearing above can be viewed as a vector average in the plane. The idea of the integral as an average should be familiar to students from earlier calculus courses, but I have not seen it introduced into the complex setting in this way. One can of course arrive at this interpretation merely by considering the formula $\int_a^b f(z(t))z'(t)\,dt$, where $z(t)$ is a parametrization of C, but ideally our argument would be introduced immediately after a general discussion of the integral as a limit of Riemann sums. This parallels the discussion of integrals of real-valued functions and forms a counterpart to the "conformality" interpretation of differentiability at a point of nonvanishing derivative, which also relies only on algebraic and geometric properties of complex numbers.

From this result it is a short step to showing that, by approximation of an arbitrary contour by polygonal line segments,

$$\int_C z\,dz = \frac{B^2 - A^2}{2}$$

for any path connecting A to B, thus exhibiting path independence.

In some cases an exact answer is not obtainable by this approach, but an approximation can be provided, giving some properties of the answer (for example, that the answer is purely real or imaginary).

Example 5. $\int_C 1/z\,dz$ where C is the straight line from $1 + i$ to $2 + 2i$: Now

$$\frac{1}{\sqrt{2}}\int_C \frac{1}{z}\,dz = \underset{z\in C}{av}\left[\frac{1}{z}\cdot\frac{1+i}{\sqrt{2}}\right],$$

so if $z = t + ti \in C$, $1 \leqslant t \leqslant 2$, $1/z = 1/t(1 + i)$; thus the average is $av_{1\leqslant t\leqslant 2}\left[1/t\sqrt{2}\right]$. Although it is not clear what the average is, we can see from this that $\int_C 1/z\,dz$ is positive and between $1/2$ and 1. [Actual value $= \log 2 = 0.693\ldots$].

Example 6. $\int_C \log z\,dz, \log z$ is principal value, C is the contour $|z| = 1$ between $-i$ and i, counterclockwise. We have $1/(\pi/2)\int_C \log z\,dz = av_{z\in C}[iz \log z] = av_{z\in C}[-z \operatorname{Arg} z]$. Considerations of the image of C under $-z \operatorname{Arg} z$ imply an estimate of si for this average, $-\pi/2 < s < 0$, thus we can estimate the value of the original integral as yi, $-\pi^2/4 < y < 0$. [Actual value is $-2i$].

Example 7. $\int_C 1\,dz, C$ any simple closed path. We have

$$\frac{1}{L(C)}\int_C 1\,dz = \underset{z\in C}{av}\, T(z),$$

and it is intuitively clear that this average is 0; thus the integral has a value of 0.

It is hoped that the estimation of integral values using this approach gives the student a feeling for the roles played by the function and contour in determining $\int_C f(z)\,dz$, can act as a check on results obtained by the more standard computations, and generally transforms the integral into an object of substance. Of course, the simplicity of the functions and contours chosen makes estimation easy; for more complicated examples computer plots of the vector fields formed by $f(z)T(z)$ for each z on C might yield graphical information which could be used for an estimate. This follows the ideas in [**1**] and [**2**], where an alternative approach to interpreting the integral involving the separation of $\int_C f(z)\,dz$ into $\int_C \overline{W}\cdot T\,ds + i\int_C \overline{W}\cdot N\,ds$, where $\overline{W} = (\operatorname{Re} f, -\operatorname{Im} f)$, and computer-aided analysis of the two *real* line integrals is discussed.

REFERENCES

1. B. Braden, Pólya's geometric picture of complex contour integrals, *Mathematics Magazine*, 60 (1987) 321–27.
2. George Pólya and Gordon Latta, Complex Variables, Wiley, 1974.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, BRUCE C. BERNDT, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, MICHAEL A. FILASETA, IRA M. GESSEL, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, JOSEPH B. MILES, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before February 29, 1992 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3451. *Proposed by M. A. Khan, Lucknow, India.*

Prove that for any positive integers m, n, N we have

$$\sum_{j=1}^{\min(n,N)} S(n,j)\frac{(N-j)^m}{(N-j)!} = \sum_{i=1}^{\min(m,N)} S(m,i)\frac{(N-i)^n}{(N-i)!},$$

where $S(n,j)$ is the Stirling number of the second kind given by

$$S(n,j) = \frac{1}{j!}\sum_{k=1}^{j}(-1)^{j-k}\binom{j}{k}k^n.$$

E 3452. *Proposed by C. A. Nicol and J. L. Selfridge, University of South Carolina, Columbia, SC.*

If n is an odd integer greater than 3 and ϕ is the Euler function, prove that there exists a prime p such that $p|(2^{\phi(n)} - 1)$ but $p \nmid n$.

E 3453. *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, MI.*

Suppose n is a positive integer greater than 2. Determine the smallest positive number c_n with the following property: Given any n distinct real numbers, there

exist two of them, say x and y, which satisfy

$$0 < \frac{x - y}{1 + xy} \leqslant c_n.$$

E 3454. *Proposed by John A. Baker, University of Waterloo, Ontario, Canada, and motivated by the work of Hans G. Kellerer, Munich, Germany.*

Suppose $f: \mathbb{R} \to \mathbb{R}$ is a function such that

$$2f(x + 1) = f(x) + f(2x)$$

for all real x.

Prove that if f is twice continuously differentiable, then f must be constant.

E 3455. *Proposed by D. G. Rogers, University of Aberdeen, Scotland, and Howard University, Washington, DC.*

It is known that if $n \equiv 0 \pmod 4$ or $n \equiv 1 \pmod 4$, then there exist permutations $(x_1, x_2, \ldots, x_n)$ of $(1, 2, \ldots, n)$ such that the differences $|x_k - k|$, $1 \leqslant k \leqslant n$, are all distinct. (Cf. E3269 [1988, 554; 1989, 843].) Prove that the number of such permutations is a multiple of 4.

E 3456. *Proposed by A. S. Cavaretta, Kent State University, Kent, OH.*

Suppose $0 = m_0 < m_1 < \cdots < m_n$ are integers such that $m_i \equiv i \pmod 2$.

(i) Prove that a real polynomial

$$c_0 + c_1 x^{m_1} + \cdots + c_n x^{m_n}, \quad \text{with } c_0 c_n \neq 0$$

has at most n real zeros, each zero being counted according to its multiplicity.

(ii) Prove that the generalized Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0^{m_1} & x_1^{m_1} & \cdots & x_n^{m_1} \\ \vdots & \vdots & \cdots & \vdots \\ x_0^{m_n} & x_1^{m_n} & \cdots & x_n^{m_n} \end{vmatrix}$$

is non-zero if $x_0, x_1, \ldots, x_n$ are any $n + 1$ distinct real numbers.

SOLUTIONS OF ELEMENTARY PROBLEMS

Feast or Famine of Friday the Thirteenths

E 3292 [1988, 873]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

(i) What is the maximum number of consecutive days with no occurrences of Friday the thirteenth?

(ii) What is the minimum number of consecutive days with four occurrences of Friday the thirteenth?

exist two of them, say x and y, which satisfy

$$0 < \frac{x - y}{1 + xy} \leqslant c_n.$$

E 3454. *Proposed by John A. Baker, University of Waterloo, Ontario, Canada, and motivated by the work of Hans G. Kellerer, Munich, Germany.*

Suppose $f: \mathbb{R} \to \mathbb{R}$ is a function such that

$$2f(x + 1) = f(x) + f(2x)$$

for all real x.

Prove that if f is twice continuously differentiable, then f must be constant.

E 3455. *Proposed by D. G. Rogers, University of Aberdeen, Scotland, and Howard University, Washington, DC.*

It is known that if $n \equiv 0 \pmod 4$ or $n \equiv 1 \pmod 4$, then there exist permutations $(x_1, x_2, \ldots, x_n)$ of $(1, 2, \ldots, n)$ such that the differences $|x_k - k|$, $1 \leqslant k \leqslant n$, are all distinct. (Cf. E3269 [1988, 554; 1989, 843].) Prove that the number of such permutations is a multiple of 4.

E 3456. *Proposed by A. S. Cavaretta, Kent State University, Kent, OH.*

Suppose $0 = m_0 < m_1 < \cdots < m_n$ are integers such that $m_i \equiv i \pmod 2$.

(i) Prove that a real polynomial

$$c_0 + c_1 x^{m_1} + \cdots + c_n x^{m_n}, \quad \text{with } c_0 c_n \neq 0$$

has at most n real zeros, each zero being counted according to its multiplicity.

(ii) Prove that the generalized Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0^{m_1} & x_1^{m_1} & \cdots & x_n^{m_1} \\ \vdots & \vdots & \cdots & \vdots \\ x_0^{m_n} & x_1^{m_n} & \cdots & x_n^{m_n} \end{vmatrix}$$

is non-zero if $x_0, x_1, \ldots, x_n$ are any $n + 1$ distinct real numbers.

SOLUTIONS OF ELEMENTARY PROBLEMS

Feast or Famine of Friday the Thirteenths

E 3292 [1988, 873]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

(i) What is the maximum number of consecutive days with no occurrences of Friday the thirteenth?

(ii) What is the minimum number of consecutive days with four occurrences of Friday the thirteenth?

To conclude the argument for (ii) we examine each of the 21 possible ordered pairs of consecutive calendar years. Such an examination reveals that no string of 14 consecutive months contains four Friday the 13ths and that strings of 15 consecutive months containing four Friday the 13ths occur only for the pairs *Cd* and *Ga*. These cases give the following 428-day periods containing four Friday the 13ths:

Case I. From September 13 of a year of type *C* through November 13 of the following year of type *d*,

Case II. From May 13 of a year of type *G* through July 13 of the following year of type *a*.

Case I occurs 14 times and Case II occurs 15 times during a 400-year period.

In each instance of Case I of (ii) the 428-day stretch containing four Friday the 13ths is immediately preceded by a 426-day stretch containing no Friday the 13ths, according to Case 1 of (i). This phenomenon occurs in the years

1714–1716, 1742–1744, 1770–1772, 1810–1812, 1838–1840, 1866–1868, 1894–1896, 1906–1908, 1934–1936, 1962–1964, 1990–1992, 2018–2020, 2046–2048, 2074–2076,

and in years differing from these by multiples of 400.

Similarly in each instance of Case II of (ii) the 428-day stretch containing four Friday the 13ths is immediately followed by a 426-day period containing no Friday the 13ths, according to Case 2 of (i). This phenomenon occurs in the years

1707–1709, 1735–1737, 1763–1765, 1791–1793, 1803–1805, 1831–1833, 1859–1861, 1887–1889, 1927–1929, 1955–1957, 1983–1985, 2011–2013, 2039–2041, 2067–2069, 2095–2097,

and in years differing from these by multiples of 400.

Finally we remark that Case 1 of (i) occurs without an immediately following instance of (ii) in the years

1725–1726, 1753–1754, 1781–1782, 1798–1799, 1821–1822, 1849–1850, 1877–1878, 1900–1901, 1917–1918, 1945–1946, 1973–1974, 2001–2002, 2029–2030, 2057–2058, 2085–2086,

and in years differing from these by multiples of 400; also Case 3 of (i) occurs in the years 1723–1724, 1751–1752, 1779–1780, 1819–1820, 1847–1848, 1875–1876, 1915–1916, 1943–1944, 1971–1972, 1999–2000, 2027–2028, 2055–2056, 2083–2084, and in years differing from these by multiples of 400.

Editorial comment. For $n \leqslant 16$ Harvey Lipman calculated the minimum number $f(n)$ of consecutive days with n occurrences of Friday the 13th and the maximum number $F(n)$ of consecutive days with n occurrences of Friday the 13th. In particular he obtained the following values:

n	0	1	2	3	4	5	6
$f(n)$		1	29	183	428	547	792
$F(n)$	426	699	944	1126	1399	1644	1735

The given problem asks for $F(0)$ and $f(4)$.

There have been at least two earlier *Monthly* problems concerned with Friday the 13th, namely E 36 [1933, 295; 1933, 607] and E 1541 [1962, 919; 1963, 759]. E 36 was devoted to B. H. Brown's famous remark that the 13th of the month is more likely to occur on Friday than on any other day of the week.

E 1541 asked for the maximum and minimum number of "Friday the 13ths" that can occur in a year. If a calendar year is meant, Bankoff's above tables provide the answers three and one respectively. If "a year" may be any 12-month period, the present problem provides the answers three and zero respectively; unfortunately the editorial comment after the solution of E 1541 erroneously stated that it is possible for a 12-month period to contain four Friday the 13ths.

Solved also by C. Aschbacher, W. D. Banks, C. W. Dodge, M. P. Eisner, B. S. Garbow, D. L. Grant, R. K. Guy, W. Janous, J. Knapp, S. Kravitz, H. Lipman, M. J. Mackinnon, A. Nijenhuis, V. Schindler, J. H. Steelman, R. Stong, M. Vowe, J. T. Ward, C. H. Webster, and S. Wolf.

The Expected Number of Selectable Subsets

E 3344 [1989, 734]. *Proposed by Richard Stanley, Massachusetts Institute of Technology, and J. Michael Steele, Princeton University.*

Suppose n points $p_1, p_2, \ldots, p_n$ are independently chosen at random in the unit square $\{(x, y)\colon 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$. Find the expected number $E(n)$ of subsets S of $N = \{p_1, p_2, \ldots, p_n\}$ for which there exists an increasing function $y = f(x)$ such that the elements of S lie below the graph of f while the elements of $N - S$ lie above the graph.

Solution by Richard Stong, University of California at Los Angeles. The expectation is $\sum_{k=0}^{n}\binom{n}{k}/k!$. Call such a set S a *selectable* set. With probability 1, we may assume the y coordinates of the points are all distinct. Hence we can modify the requirement for selectability so that the selected set lies on or below the function. If $S = \{p_i = (x_i, y_i)\}$ is selectable, then S is selectable by the step-function $f_S(x) = \sup\{y_i\colon x_i \leqslant x\}$. Let $T = \tau(S)$ be the minimal subset of S for which $f_T = f_S$; this is unique and consists of the points of S that are "upper leftmost" in the region of the square on and below f_S. Note that T forms a chain in the usual chain-product partial ordering on the points in the unit square.

Conversely, for any subset $T \subseteq N$ that forms a chain in this ordering, we can define $S = \sigma(T)$ to be the set of points in N on or below f_T and observe that f_T selects this set. Furthermore, $S = \sigma(\tau(S))$, so σ and τ establish a bijection between the selectable subsets and the subsets that form chains in the coordinate ordering. Therefore, the expectation can be rewritten as the sum over all subsets of indices $M \subseteq \{1, \ldots, n\}$ of the probability that the points indexed by M form a chain in the coordinate ordering. This probability is simply $1/|M|!$.

Editorial comment. Richard K. Guy noted that some related results are given in Louis W. Shapiro, A lattice path lemma and an application in enzyme kinetics, *J. Statistical Planning and Inference* 14 (1986), 115–122, and in Roger B. Eggleton and Richard K. Guy, Catalan strikes again! How likely is a function to be convex?, *Math. Magazine* 61 (1988), 211–219.

Solved also by M. Bazant (student), D. Callan, C. Schoen, A. J. Schwenk, and the proposers.

Suppose f and g are nonnegative continuous functions compactly supported on the real line. Hölder's inequality gives

$$\sup_x \int f(x-y)g(y)\,dy \leqslant \|f\|_p \|g\|_q$$

for positive numbers p, q with $p^{-1} + q^{-1} = 1$. Prove or disprove that

$$\int \sup_y \{f(x-y)g(y)\}\,dx \geqslant \|f\|_p \|g\|_q.$$

Solution I by Philip Spencer and Richard Stong, Harvard University, Cambridge, MA. The stated inequality does hold. To see this, let y_0 be any point at which g attains its maximum $\|g\|_\infty$. Then $\sup_y\{f(x-y)g(y)\} \geqslant f(x-y_0)\|g\|_\infty$ and thus $\int \sup_y\{f(x-y)g(y)\}\,dx \geqslant \|f\|_1\|g\|_\infty$. Similarly, $\int \sup_y\{f(x-y)g(y)\}\,dx \geqslant \|f\|_\infty\|g\|_1$. Combining these two bounds gives

$$\int \sup_y \{f(x-y)g(y)\}\,dx \geqslant \|f\|_1^{1/p}\|g\|_\infty^{1/p}\|f\|_\infty^{1/q}\|g\|_1^{1/q}.$$

But

$$\|f\|_1^{1/p}\|f\|_\infty^{1/q} = \left[\|f\|_\infty^{p-1}\int f(x)\,dx\right]^{1/p} \geqslant \left[\int f(x)^p\,dx\right]^{1/p} = \|f\|_p$$

and similarly $\|g\|_1^{1/q}\|g\|_\infty^{1/p} \geqslant \|g\|_q$, from which the desired inequality follows.

Solution II by Andrew Kresch, student, Yale University, New Haven, CT. The assertion is true. Applying Hölder's inequality to the functions g^{q-1} and f^{p-1} yields

$$\|g^{q-1}\|_p\|f^{p-1}\|_q \geqslant \int (f(x-y))^{p-1}(g(y))^{q-1}\,dy$$

for any x. This may be rewritten as

$$\|f\|_p^{p-1}\|g\|_q^{q-1} \geqslant \int (f(x-y))^{p-1}(g(y))^{q-1}\,dy,$$

from which it follows that

$$\sup_y \{f(x-y)g(y)\}\|f\|_p^{p-1}\|g\|_q^{q-1}$$

$$\geqslant \int f(x-y)g(y)(f(x-y))^{p-1}(g(y))^{q-1}\,dy.$$

Integrating both sides with respect to x and making the substitution $z = x - y$ on the right yields

$$\|f\|_p^{p-1}\|g\|_q^{q-1}\int \sup_y \{f(x-y)g(y)\}\,dx$$

$$\geqslant \int (f(z))^p\,dz \int (g(y))^q\,dy = \|f\|_p^p\|g\|_q^q,$$

from which the desired inequality follows directly.

Editorial note. A stronger result is proved by L. Leindler (On a certain converse of Hölder's inequality, *Acta Szeged*, 33 (1972) 217–223), who showed that

$$\int \sup_y \{f(x-y)g(y)\}\,dx \geqslant p^{1/p}q^{1/q}\|f\|_p\|g\|_q$$

for any positive p and q with $p^{-1} + q^{-1} = 1$. Leindler's result is best possible in the sense that equality holds if f is the characteristic function of $[0, 1/p]$ and g is the characteristic function of $[0, 1/q]$.

Solved also by A. Riese and the proposer.

A Sum of Permutation Averages

E 3371 [1990, 151]. *Proposed by A. M. Garsia and B. A. Sethuraman, University of California at San Diego, La Jolla, CA.*

Let S_n denote the symmetric group on the n symbols $1, 2, \ldots, n$. Given σ in S_n, let

$$M(\sigma) = \{i: 1 \leqslant i \leqslant n, i \geqslant \sigma(i)\}.$$

Put

$$P(\sigma) = |M(\sigma)|^{-1} \sum_{i \in M(\sigma)} \{i + \sigma(i)\}.$$

Prove that

$$\sum_{\sigma \in S_n} P(\sigma) = (n+1)!.$$

Solution by Ron Adin, The Hebrew University, Jerusalem, Israel. For each $\sigma \in S_n$, define $\sigma^* \in S_n$ by $\sigma^*(i) = j$ if and only if $\sigma(n+1-j) = n+1-i$. The map $\sigma \to \sigma^*$ is an involution on S_n. Since $i \geqslant j$ if and only if $n+1-j \geqslant n+1-i$, we obtain $|M(\sigma^*)| = |M(\sigma)|$, and hence

$$\begin{aligned} P(\sigma) + P(\sigma^*) &= |M(\sigma)|^{-1} \sum_{i \in M(\sigma)} [i + \sigma(i) + (n+1-\sigma(i)) + n + 1 - i] \\ &= 2(n+1). \end{aligned}$$

Therefore, $2\Sigma P(\sigma) = \Sigma[P(\sigma) + P(\sigma^*)] = \Sigma_\sigma 2(n+1) = 2(n+1)!$.

Solved also by P. Bloomsburg, P. Brosnan (student), D. Callan, R. J. Chapman, T. H. Crocker, B. M. M. de Weger (Netherlands), K. L. Dove, J. L. Drost, A. A. Jagers (Netherlands), A. Kresch (student), R. Lipsett, J. H. Nieto (Venezuela), J. H. Steelman, R. Stong, D. Wolfe, Central Michigan University Problem Group, National Security Agency Problems Group, and the proposers.

The Sum of Arctangents of Reciprocal Squares

E 3375 [1990, 239]. *Proposed by R. J. Chapman, Merton College, Oxford, U.K.*

Express

$$\tan\left\{\sum_{n=1}^{\infty} \arctan(1/n^2)\right\}$$

Editorial note. A stronger result is proved by L. Leindler (On a certain converse of Hölder's inequality, *Acta Szeged*, 33 (1972) 217–223), who showed that

$$\int \sup_y \{f(x-y)g(y)\}\,dx \geqslant p^{1/p}q^{1/q}\|f\|_p\|g\|_q$$

for any positive p and q with $p^{-1} + q^{-1} = 1$. Leindler's result is best possible in the sense that equality holds if f is the characteristic function of $[0, 1/p]$ and g is the characteristic function of $[0, 1/q]$.

Solved also by A. Riese and the proposer.

A Sum of Permutation Averages

E 3371 [1990, 151]. *Proposed by A. M. Garsia and B. A. Sethuraman, University of California at San Diego, La Jolla, CA.*

Let S_n denote the symmetric group on the n symbols $1, 2, \ldots, n$. Given σ in S_n, let

$$M(\sigma) = \{i: 1 \leqslant i \leqslant n, i \geqslant \sigma(i)\}.$$

Put

$$P(\sigma) = |M(\sigma)|^{-1} \sum_{i \in M(\sigma)} \{i + \sigma(i)\}.$$

Prove that

$$\sum_{\sigma \in S_n} P(\sigma) = (n+1)!.$$

Solution by Ron Adin, The Hebrew University, Jerusalem, Israel. For each $\sigma \in S_n$, define $\sigma^* \in S_n$ by $\sigma^*(i) = j$ if and only if $\sigma(n+1-j) = n+1-i$. The map $\sigma \to \sigma^*$ is an involution on S_n. Since $i \geqslant j$ if and only if $n+1-j \geqslant n+1-i$, we obtain $|M(\sigma^*)| = |M(\sigma)|$, and hence

$$\begin{aligned} P(\sigma) + P(\sigma^*) &= |M(\sigma)|^{-1} \sum_{i \in M(\sigma)} [i + \sigma(i) + (n+1-\sigma(i)) + n + 1 - i] \\ &= 2(n+1). \end{aligned}$$

Therefore, $2\Sigma P(\sigma) = \Sigma[P(\sigma) + P(\sigma^*)] = \Sigma_\sigma 2(n+1) = 2(n+1)!$.

Solved also by P. Bloomsburg, P. Brosnan (student), D. Callan, R. J. Chapman, T. H. Crocker, B. M. M. de Weger (Netherlands), K. L. Dove, J. L. Drost, A. A. Jagers (Netherlands), A. Kresch (student), R. Lipsett, J. H. Nieto (Venezuela), J. H. Steelman, R. Stong, D. Wolfe, Central Michigan University Problem Group, National Security Agency Problems Group, and the proposers.

The Sum of Arctangents of Reciprocal Squares

E 3375 [1990, 239]. *Proposed by R. J. Chapman, Merton College, Oxford, U.K.*

Express

$$\tan\left\{\sum_{n=1}^{\infty} \arctan(1/n^2)\right\}$$

in closed form (in terms of a finite number of trigonometric and hyperbolic functions).

Solution by Amites Sarkar, Winchester College, Hampshire, Great Britain.

$$\begin{aligned}
\tan\left\{\sum_{n=1}^{\infty}\arctan(1/n^2)\right\} &= \tan\left\{\sum_{n=1}^{\infty}\operatorname{Arg}\left(1+\frac{i}{n^2}\right)\right\}\\
&= \tan\left\{\operatorname{Arg}\left(\prod_{n=1}^{\infty}\left(1+\frac{\left(\pi(1+i)/\sqrt{2}\right)^2}{n^2\pi^2}\right)\right)\right\}\\
&= \tan\left\{\operatorname{Arg}\left(\frac{\sinh\left(\pi(1+i)/\sqrt{2}\right)}{\pi(1+i)/\sqrt{2}}\right)\right\}\\
&= \tan\left\{\arctan\left(\frac{\tan(\pi/\sqrt{2})-\tanh(\pi/\sqrt{2})}{\tan(\pi/\sqrt{2})+\tanh(\pi/\sqrt{2})}\right)\right\}\\
&= \frac{\tan(\pi/\sqrt{2})-\tanh(\pi/\sqrt{2})}{\tan(\pi/\sqrt{2})+\tanh(\pi/\sqrt{2})}.
\end{aligned}$$

The numerical value of this quantity is about 6.798.

Editorial note. The problem and its generalizations have made frequent earlier appearances dating back to J. W. L. Glaisher in 1878 and most recently as Problem 399 of *The College Mathematics Journal*. Readers cited previous appearances in works of S. L. Loney, S. Ramanujan, T. J. I. Bromwich, E. W. Hobson, L. B. W. Jolley, M. S. Klamkin & M. L. Glasser, and N. Lord. For references and a more complete history, see *The College Mathematics Journal*, vol. 21, No. 3, May 1990, pp. 253–4.

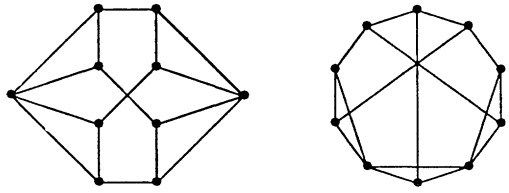
Solved also by 38 others and the proposer.

Large Graphs with No 4-cycle

E 3387 [1990, 428]. *Proposed by Douglas B. West, University of Illinois, Urbana, IL.*

Determine the maximum possible number of edges in a simple graph on ten vertices that contains no 4-cycle, and find all graphs achieving this maximum.

Solution by Arie Bialostocki, University of Idaho, Moscow, Idaho, and J. Schönheim, Tel Aviv University, Israel, essentially as given in reference [**1**]. The maximum number of edges is 16, achieved by the two graphs drawn below.



Suppose G is a 10-vertex graph with no 4-cycle. Let v be a vertex of maximum degree, S its set of neighbors, and T its set of non-neighbors. Suppose $|S| = k$, so that $|T| = 9 - k$. No vertex of S or T can have two neighbors in S, so there are at most $\lfloor k/2 \rfloor$ edges within S and at most $9 - k$ edges between S and T. The remaining edges are within T. For $k = 4, 5, 6, 7, 8, 9$, the maximum number of edges in T with no 4-cycle is $6, 4, 3, 1, 0, 0$, respectively. Including edges incident with v, the number of edges in G is bounded by $17, 15, 15, 13, 13, 13$ for $k = 4, 5, 6, 7, 8, 9$, respectively. Since maximum degree 3 implies G has at most 15 edges, G must have maximum degree exactly 4 to exceed 15 edges.

Consider $k = 4$. We claim that G cannot have 17 edges, because this requires 6 edges in T, achievable only by a vertex of degree 4 and two non-incident edges, which prevents placing 5 edges between S and T. Moreover, 16 edges are achievable only by 6 edges in T, 2 in S, and 4 between S and T (the first graph above), or by 5 edges in T, 2 in S, and 5 between S and T (the second graph above). For the first graph, placing one edge between S and T readily forces the others. For the second graph, the only way to place 5 edges in T and 5 between S and T is for the vertices in T to induce a 5-cycle, after which the remaining edges are again readily forced.

Editorial comment. Let $f(n)$ be the maximum number of edges in an n-vertex graph with no 4-cycle. The asymptotic behavior of f has been well studied. This is related to the problem of Zarankiewicz, and a thorough discussion can be found in [2]; for large n we have $(n/2)(n^{1/2} - n^{1/3}) \leqslant f(n) \leqslant (n/2)((n - 3/4)^{1/2} + 1/2)$. The constructions, provided in 1966 by Erdös-Renyi-Sós [5] and by W. G. Brown [3], use projective planes, available when $n = q^2 + q + 1$, where q is a prime power. Recently, Zoltán Füredi [6] has proved that these are best possible for $n = q^2 + q + 1$ and q a prime power exceeding 13, where $f(n) = \frac{1}{2}q(q + 1)^2$; he also characterized the extremal graphs.

The editors were not aware that the exact answers for small n have also been determined. Bialostocki and Schönheim determined $f(10)$ and $f(11)$. A determination of $f(n)$ for $n \leqslant 21$, along with all the extremal graphs, was apparently made by William McCuaig in 1984 but never published. These results were also obtained by C. R. J. Clapham, A. Flockhart, and J. Sheehan [4]. Letting $g(n)$ be the number of extremal graphs, the results are

n	4	5	6	7	8	9	10	11	12
$f(n)$	4	6	7	9	11	13	16	18	21
$g(n)$	1	1	4	5	5	10	2	11	3

n	13	14	15	16	17	18	19	20	21
$f(n)$	24	27	30	33	36	39	42	46	50
$g(n)$	2	1	2	2	1	1	5	1	1

REFERENCES

1. A. Bialostocki and J. Schönheim, On some Turán and Ramsey numbers for C_4, in Graph Theory and Combinatorics, Academic Press, 1984, pp. 29–33.
2. B. Bollobás, Extremal Graph Theory, Academic Press, 1978, pp. 312–314.
3. W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.*, 9 (1966) 281–285.

4. C. R. J. Clapham, A. Flockhart, and J. Sheehan, Graphs Without Four-Cycles, *J. Graph Theory*, 13 (1989) 29–47.
5. P. Erdös, A. Rényi, and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.*, 1 (1966) 215–235.
6. Z. Füredi, Quadrilateral-free graphs with maximum number of edges, to appear.

In addition to the references above, solutions were submitted by R. J. Chapman (Great Britain), M. E. Kuczma (Poland), C. Rousseau, A. J. Schwenk, R. Stong, and the proposer.

ADVANCED PROBLEMS

6664. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

Let G be a graph whose edges can be covered by n complete subgraphs with n vertices each (i.e., G is the union of n copies of K_n, with no restriction on shared vertices).

(a) Prove that the chromatic number of G is less than $1 + n\sqrt{n-1}$.

(b) Prove that this bound is asymptotically best possible, i.e., if $f(n)$ is the maximum chromatic number of a graph constructed in this way, then $f(n) = \{1 + o(1)\}n^{3/2}$.

6665. *Proposed by José Luis Palacios, New Jersey Institute of Technology, Newark, NJ, and Dennis P. Sandell, Swedish University of Agricultural Sciences, Garpenberg, Sweden.*

Let S_n $(n \geqslant 0)$ be a simple symmetric random walk, i.e., $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$ for $n > 0$, where the X_i are independent identically distributed random variables with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Let N be an arbitrary positive integer and let T be the first time that the difference between the maximum and minimum of the random walk is N, i.e., let

$$T = \min\Big\{n\colon \max_{0\leqslant k\leqslant n} S_k - \min_{0\leqslant k\leqslant n} S_k = N\Big\}.$$

Find the expected value of T.

6666. *Proposed by Marcel F. Neuts, University of Arizona, Tucson, AZ.*

Suppose A and B are n by n matrices with positive real entries, suppose A is non-singular, and let α denote the largest eigenvalue of BA.

(a) Prove that there exists an n by n matrix X with positive entries satisfying

$$X = A + XBX \qquad (*)$$

if and only if $\alpha \leqslant \frac{1}{4}$.

(b) If $\alpha \leqslant \frac{1}{4}$, prove that

(i) there is a solution X_0 of (*) with positive entries such that if X_1 is any other solution of (*) with positive entries, then all the entries of $X_1 - X_0$ are non-negative;

(ii) the maximum eigenvalue of BX_0 is $(1 - \sqrt{1 - 4\alpha})/2$;

(iii) if X_1 is any solution of (*) with positive entries other than the minimal solution X_0, then the maximum eigenvalue of BX_1 is $(1 + \sqrt{1 - 4\alpha})/2$.

4. C. R. J. Clapham, A. Flockhart, and J. Sheehan, Graphs Without Four-Cycles, *J. Graph Theory*, 13 (1989) 29–47.
5. P. Erdös, A. Rényi, and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.*, 1 (1966) 215–235.
6. Z. Füredi, Quadrilateral-free graphs with maximum number of edges, to appear.

In addition to the references above, solutions were submitted by R. J. Chapman (Great Britain), M. E. Kuczma (Poland), C. Rousseau, A. J. Schwenk, R. Stong, and the proposer.

ADVANCED PROBLEMS

6664. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

Let G be a graph whose edges can be covered by n complete subgraphs with n vertices each (i.e., G is the union of n copies of K_n, with no restriction on shared vertices).

(a) Prove that the chromatic number of G is less than $1 + n\sqrt{n-1}$.

(b) Prove that this bound is asymptotically best possible, i.e., if $f(n)$ is the maximum chromatic number of a graph constructed in this way, then $f(n) = \{1 + o(1)\}n^{3/2}$.

6665. *Proposed by José Luis Palacios, New Jersey Institute of Technology, Newark, NJ, and Dennis P. Sandell, Swedish University of Agricultural Sciences, Garpenberg, Sweden.*

Let S_n $(n \geqslant 0)$ be a simple symmetric random walk, i.e., $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$ for $n > 0$, where the X_i are independent identically distributed random variables with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Let N be an arbitrary positive integer and let T be the first time that the difference between the maximum and minimum of the random walk is N, i.e., let

$$T = \min\left\{n: \max_{0 \leqslant k \leqslant n} S_k - \min_{0 \leqslant k \leqslant n} S_k = N\right\}.$$

Find the expected value of T.

6666. *Proposed by Marcel F. Neuts, University of Arizona, Tucson, AZ.*

Suppose A and B are n by n matrices with positive real entries, suppose A is non-singular, and let α denote the largest eigenvalue of BA.

(a) Prove that there exists an n by n matrix X with positive entries satisfying

$$X = A + XBX \qquad (*)$$

if and only if $\alpha \leqslant \frac{1}{4}$.

(b) If $\alpha \leqslant \frac{1}{4}$, prove that

(i) there is a solution X_0 of (*) with positive entries such that if X_1 is any other solution of (*) with positive entries, then all the entries of $X_1 - X_0$ are non-negative;

(ii) the maximum eigenvalue of BX_0 is $(1 - \sqrt{1 - 4\alpha})/2$;

(iii) if X_1 is any solution of (*) with positive entries other than the minimal solution X_0, then the maximum eigenvalue of BX_1 is $(1 + \sqrt{1 - 4\alpha})/2$.

SOLUTIONS OF ADVANCED PROBLEMS

Inscriptible Polygons

6571 [1988, 359]. *Proposed by Glenn Ierley, Michigan Technological University, Houghton, MI.*

(a) Let $A(n)$ be the maximum area of a polygon with n sides of lengths $1, 2, \ldots, n$, where $n \geqslant 4$. It is known that the maximum area occurs for a polygon inscribed in a circle. (Cf. G. Pólya, Mathematics and Plausible Reasoning, Volume 1, Princeton, 1954, pp. 174–177.) Let $B(n)$ be the area of a regular polygon with n sides and perimeter $1 + 2 + \cdots + n$. Prove that

$$1 - \frac{A(n)}{B(n)} \sim \frac{\pi^2}{3n^2} \qquad (n \to \infty).$$

(b) For $1/2 < q < 1$ let $A(q, n)$ be the maximum area of a polygon with n sides of lengths $1, q, q^2, \ldots, q^{n-1}$, respectively, where n is large enough so that $q + q^2 + \cdots + q^{n-1} > 1$. Let $B(q, n)$ be the area of a regular polygon with n sides and perimeter $1 + q + q^2 + \cdots + q^{n-1}$. Prove that $c(q) = \lim_{n\to\infty} A(q, n)/B(q, n)$ exists and find

$$\lim_{q\to 1^-} \frac{1 - c(q)}{(1 - q)^2}$$

Solution by William A. Newcomb, Lawrence Livermore Laboratory, Livermore, CA. The limit for (*b*) is $\pi^2/9$. We shall consider the more general problem of maximizing the area of an n-gon with sides of prescribed positive lengths $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_n$ satisfying $l_1 < l_2 + \cdots + l_n$. Steiner's Theorem, cited by Pólya, states that this maximum is achieved by such an n-gon when its vertices lie on a circle. We shall first prove that the radius of such a circle is uniquely determined.

Let $F(l, r) = \sin^{-1}(l/(2r))$. Observe that a chord of length l divides a circle of radius $r \geqslant l/2$ into minor and major arcs of angular measure $2F(l, r)$ and $2\pi - 2F(l, r)$. If a polygon is inscribed in a circle, then at most one side, the largest, might subtend a major arc. Accordingly, in our case, the sides of length $l_2, \ldots, l_n$ must cut off minor arcs of total angular measure

$$2F(r) = \sum_{j=2}^{n} 2F(l_j, r) = \sum_{j=2}^{n} 2\sin^{-1}\big(l_j/(2r)\big). \tag{1}$$

Now $F(r) < \pi$, and of course $r \geqslant l_2/2$. Link chords of length $l_2, \ldots, l_n$ (in any order) so that their vertices lie on a circle of radius r. Then the distance between the endpoints is

$$G(r) = 2r \sin F(r), \tag{2}$$

and r is a possible radius for a circumscribing circle precisely when $G(r) = l_1$, $F(r) < \pi$ and $r \geqslant l_1/2$.

Observe from (1) that F is a strictly decreasing function on $[l_2/2, \infty) \supseteq [l_1/2, \infty)$. We define s as follows: if $F(l_1/2) \leqslant \pi$ then $s = l_1/2$; if $F(l_1/2) > \pi$ then s is the unique value in $[l_1/2, \infty)$ such that $F(s) = \pi$. We now show that $G(r)$ is strictly increasing on $[s, \infty)$ and that the equation $G(r) = l_1$ has exactly one solution there.

First, some reduction yields

$$G'(r) = 2\sin F(r) - 2\cos(F(r)) \sum_{j=2}^{n} \tan\big(F(l_j, r)\big). \tag{3}$$

If $F(r) \in [0, \pi/2]$, then the superadditivity of the tangent implies that $G'(r) \geqslant 0$; if $F(r) \in (\pi/2, \pi)$ the same conclusion is immediate. Finally, it is easy to see that

$$\lim_{r\to\infty} G(r) = l_2 + \cdots + l_n,$$

a sum that exceeds l_1 by hypothesis. If $s = l_1/2$, then $F(s) = F(l_1/2) \leqslant \pi$, so $G(s) \leqslant 2s = l_1$. If $s > l_1/2$, then $F(s) = \pi$, and hence $G(s) = 0$. In either case $G(s) \leqslant l_1$, so the equation $G(r) = l_1$ has exactly one solution in the allowable range, and so the circle's radius is uniquely determined.

We turn now to a more qualitative description. Suppose a polygon with sides l_i as above is inscribed in a circle of radius r. If the arc subtended by l_1 is minor, then by adding the interior angles we find that

$$2\pi = \sum_{j=1}^{n} 2F(l_j, r). \tag{4}$$

Since the right-hand side of (4) is decreasing in r on $[l_1/2, \infty)$, it follows that this case is equivalent to

$$2\pi \leqslant \sum_{j=1}^{n} 2F(l_j, l_1/2) = \pi + 2\sum_{j=2}^{n} \sin^{-1}(l_j/l_1). \tag{5}$$

In case (5) does not hold, l_1 subtends a major arc and

$$2\pi = 2\pi - 2F(l_1, r) + \sum_{j=2}^{n} 2F(l_j, r). \tag{6}$$

For the two particular problems posed it will turn out that we need only deal in detail with the first case.

Let us now consider (a). We have $l_1 = n, \ldots, l_n = 1$ and

$$\sum_{j=2}^{n} \sin^{-1}(l_j/l_1) \geqslant \sin^{-1}\left(\tfrac{1}{4}\right) + \sin^{-1}\left(\tfrac{1}{2}\right) + \sin^{-1}\left(\tfrac{3}{4}\right) \doteq .517\pi, \tag{7}$$

so that by (5) we are in the first case. Let $R = R(n)$ be the radius of the inscribing circle. Then, by (4),

$$\pi = \sum_{j=1}^{n} \sin^{-1}(j/(2R)). \tag{8}$$

Since for $0 \leqslant \alpha \leqslant 1$, we have

$$\alpha \leqslant \sin^{-1}\alpha \leqslant \frac{\pi}{2}\alpha,$$

equality (8) implies that $C_2 n^2 \leqslant R(n) \leqslant C_1 n^2$, so $j/(2R) = O(n^{-1})$. From the Taylor expansion

$$\sin^{-1}\alpha = \alpha + \alpha^3/6 + O(\alpha^5),$$

we obtain from (8) the expansion

$$\pi = \frac{n(n+1)}{4R} + \frac{n^2(n+1)^2}{192R^3} + O(n^{-4}), \tag{9}$$

which easily inverts to

$$R = \frac{n(n+1)}{4\pi} + \frac{\pi}{12} + O(n^{-1}). \tag{10}$$

Furthermore,

$$\begin{aligned} A(n) &= \sum_{j=1}^{n} \frac{1}{2} j\{R^2 - (j/2)^2\}^{1/2} = \frac{R}{2} \sum_{j=1}^{n} j\left(1 - (j/(2R))^2\right)^{1/2} \\ &= \frac{R}{2} \sum_{j=1}^{n} \left[j - \frac{1}{8} \frac{j^3}{R^2} + O\left(\frac{j^5}{R^4}\right) \right] \\ &= \frac{n(n+1)}{4} R - \frac{1}{64} \frac{n^2(n+1)^2}{R} + O(1) \\ &= \frac{n^2(n+1)^2}{16\pi} \left(1 - \frac{2\pi^2}{3n^2} + O(n^{-3})\right). \end{aligned} \tag{11}$$

On the other hand, it is immediate that

$$B(n) = \frac{n(n+1)^2}{16} \cot \frac{\pi}{n} = \frac{n^2(n+1)^2}{16\pi} \left(1 - \frac{\pi^2}{3n^2} + O(n^{-4})\right), \tag{12}$$

from which the desired result follows.

For (*b*) we prove a stronger result, that $A(q) = \lim_{n\to\infty} A(q, n)$ and $B(q) = \lim_{n\to\infty} B(q, n)$ each exist. Indeed, $B(q, n)$ is the area of a regular n-gon with perimeter $(1 - q^n)/(1 - q)$, so $B(q)$ is the area of the circle with circumference $(1 - q)^{-1}$, namely $(4\pi)^{-1}(1 - q)^{-2}$. We shall prove the existence of the first limit by using the monotone convergence theorem.

Suppose P is the maximal (convex) n-gon with sides $l_1 \geqslant \cdots \geqslant l_n$ and suppose $l < 2l_n$. We claim that the maximal area of the $(n + 1)$-gon with sides $l_1, \ldots, l_n$, l is greater than the area of P. Indeed, given P, simply attach an isosceles triangle with sides l_n, l_n, and l onto the side of P with length l_n. This gives a particular $(n + 1)$-gon with sides $l_1, \ldots, l_n$, l with area greater than the area of P; the maximal area is even larger. (That some restriction on l is needed for monotonicity can be seen by examining the quadrilateral with sides 1, 1, 1, and 2.99.)

Since $q^n < q^{n-1} < 2q^{n-1}$, it follows that $A(q, n) < A(q, n + 1)$. Since the sequence $\{A(q, n)\}$ is uniformly bounded by $B(q)$, we see that $\lim_{n\to\infty} A(q, n)$ exists.

If $\sum_{j=1}^{n-1} \sin^{-1}(q^j) \leqslant \pi/2$, then we are in the second case. However, if q is sufficiently close to 1, this will not happen (in fact,

$$\sum_{j=1}^{\infty} \sin^{-1} q^j > \frac{\pi}{2}$$

for $q > .602527\ldots$). We now write $q = 1 - \varepsilon$ where ε is sufficiently small. We

already know that $B(q) = (4\pi\varepsilon^2)^{-1}$. To find $A(q)$ we proceed as before. The radius $R_n = R(q, n)$ of the n-gon with sides $1, q, \ldots, q^{n-1}$ is determined by

$$\pi = \sum_{j=0}^{n-1} \sin^{-1}\left(q^j/(2R_n)\right) \tag{13}$$

and (cf. (11))

$$A(q, n) = \sum_{j=0}^{n-1} \frac{R_n}{2} q^j \left(1 - \frac{q^{2j}}{(2R_n)^2}\right)^{1/2}. \tag{14}$$

Since everything converges uniformly, define R by

$$\pi = \sum_{j=0}^{\infty} \sin^{-1}\left(\frac{q^j}{2R}\right), \tag{15}$$

so that

$$A(q) = \frac{R}{2} \sum_{j=0}^{\infty} q^j \left(1 - \frac{q^{2j}}{4R^2}\right)^{1/2}. \tag{16}$$

The previous estimate of $\sin^{-1}\alpha$, together with (15), shows that R has order $(1 - q)^{-1} = \varepsilon^{-1}$. Again, by Taylor series estimates,

$$\pi = \frac{1}{2R\varepsilon} + \frac{1}{144R^3\varepsilon} + O(\varepsilon^3). \tag{17}$$

Again this may be inverted, in this case to

$$R = \frac{1}{2\pi\varepsilon} + \frac{\pi\varepsilon}{36} + O(\varepsilon^2). \tag{18}$$

A similar calculation based on (16) shows that

$$\begin{aligned} A(q) &= \frac{R}{2} \sum_{j=0}^{\infty} \left(q^j - \frac{q^{3j}}{8R^2} + O\left(\frac{q^{5j}}{R^4}\right)\right) \\ &= \frac{R}{2\varepsilon} - \frac{1}{16R\varepsilon(1 + q + q^2)} + O(\varepsilon^2) \\ &= \frac{1}{4\pi\varepsilon^2} - \frac{\pi}{36} + O(\varepsilon). \end{aligned} \tag{19}$$

Finally we have

$$c(q) = \frac{A(q)}{B(q)} = 1 - \frac{\pi^2}{9}\varepsilon^2 + O(\varepsilon^3), \tag{20}$$

so the desired limit is $\pi^2/9$.

Editorial comment. Both Newcomb and the proposer also proved that

$$\lim_{\delta \to 0} \frac{A\left(\frac{1}{2} + \delta\right)}{\sqrt{\delta}} = \frac{2}{\sqrt{7}}.$$

Solved also by the proposer.

when $s \neq \bar{s}$, for s in a reduced residue system modulo c. Thus, in view of the multiplicativity of χ, we have

$$\det M = \chi\left(\prod_{k=1}^{c}{}^{*} k\right)(-1)^{\{\phi(c)-N\}/2} \det A_c^2,$$

where $N = N_c$ is the number of solutions of $x^2 \equiv 1 \pmod{c}$. Now let $\delta_c = -1$ if c has a primitive root (i.e., if c is either 4, a power of an odd prime, or twice a power of an odd prime), and let $\delta_c = 1$ otherwise. Then by Gauss's generalization of Wilson's theorem (Theorem 11.6 or Theorem 12.17 of [**2**]) we have

$$\prod_{k=1}^{c}{}^{*} k = \delta_c.$$

Also $(-1)^{N/2} = \delta_c$ (Theorem 11.5 of [2]). Thus

$$\det M = \delta_c \chi(\delta_c)(-1)^{\phi(c)/2} \det A_c^2.$$

Accordingly, in order to establish the results claimed above, it suffices to prove the following result.

LEMMA. *If c is a squarefree number exceeding 2, then*

$$\det A_c^2 = (-1)^{\phi(c)/2} \prod_{p|c} p^{(p-2)\phi(c/p)} \qquad (*)$$

where the product is taken over the primes p dividing c. If c is a positive integer divisible by the square of some prime, then $\det A_c^2 = 0$.

To prove the lemma we note that

$$\left(A_c^2\right)_{jk} = \sum_{t=1}^{c}{}^{*} \varepsilon^{t(j+k)} = C_c(j+k),$$

which is Ramanujan's sum and has the evaluation

$$C_c(j+k) = \sum_{d|(c,\, j+k)} d\mu(c/d).$$

(Cf. [**1**] p. 160.)

Consider first the case when c is squarefree. We proceed by induction on the number of prime factors of c. The case when c has just one odd prime factor was covered in the solution of part (a). We suppose $c = pm$, where p is a prime, $p \nmid m$, and $m > 2$.

If $p = 2$, i.e., if $c = 2m$, with m odd, then $\phi(c) = \phi(m)$ and a reduced residue system modulo c is also a reduced residue system modulo m. By the multiplicativity of Ramanujan's sum ([**1**], p. 163) we have

$$\left(A_c^2\right)_{jk} = C_{2m}(j+k) = C_2(j+k)C_m(j+k) = (-1)^{j+k} C_m(j+k).$$

Since j and k are both odd

$$\left(A_c^2\right)_{jk} = C_m(j+k) = \left(A_m^2\right)_{jk}.$$

Thus if $(*)$ holds for m, it also holds for c.

Suppose p is an odd prime, $c = pm$, $p \nmid m$, and $m > 2$. Let S be a reduced residue system modulo c. Then ([**1**]), p. 125) S can be partitioned into $p - 1$

distinct sets $S_1, \ldots, S_{p-1}$, where the elements of S_i are congruent to i modulo p and form a reduced residue system modulo m. We arrange the sets in the order $S_1, \ldots, S_{p-1}$; within each set, we put the elements in increasing order (from 1 to $m-1$) of congruence class modulo m. With this ordering the elements of S may be labeled $a_1, a_2, \ldots, a_{\phi(c)}$. If d_{jk} is the j, k entry of A_c^2, we have

$$d_{jk} = \sum_{t=1}^{c}{}^{*} \varepsilon^{t(a_j + a_k)} = C_c(a_j + a_k).$$

From the multiplicativity of Ramanujan's sum we have

$$d_{jk} = C_p(a_j + a_k) \cdot C_m(a_j + a_k).$$

Since

$$C_p(a_j + a_k) = -1 \quad \text{if } a_j + a_k \not\equiv 0 \pmod{p}$$

and

$$C_p(a_j + a_k) = p - 1 \quad \text{if } a_j + a_k \equiv 0 \pmod{p},$$

we find that A_c^2 has the form

$$\begin{bmatrix} -A_m^2 & -A_m^2 & \cdots & -A_m^2 & (p-1)A_m^2 \\ -A_m^2 & -A_m^2 & \cdots & (p-1)A_m^2 & -A_m^2 \\ \cdots & & \cdots & & \cdots \\ -A_m^2 & (p-1)A_m^2 & \cdots & -A_m^2 & -A_m^2 \\ (p-1)A_m^2 & -A_m^2 & \cdots & -A_m^2 & -A_m^2 \end{bmatrix}.$$

Performing the same operations as in the solution of (a) except that we now apply them to square blocks of size $\phi(m)$ instead of to numbers, we find that A_c^2 has the same determinant as

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & A_m^2 \\ 0 & 0 & \cdots & pA_m^2 & 0 \\ & \cdots & \cdots & & \cdots \\ 0 & pA_m^2 & \cdots & 0 & 0 \\ pA_m^2 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Now the number of rows of this matrix is $(p-1)\phi(m) \equiv 0 \pmod{4}$ and clearly $(p-1)\phi(m)/2$ row exchanges bring it to the block-diagonal form

$$\begin{bmatrix} pA_m^2 & 0 & \cdots & 0 & 0 \\ 0 & pA_m^2 & \cdots & 0 & 0 \\ & \cdots & \cdots & & \cdots \\ 0 & 0 & \cdots & pA_m^2 & 0 \\ 0 & 0 & \cdots & 0 & A_m^2 \end{bmatrix}.$$

Hence $\det A_c^2 = p^{(p-2)\phi(m)}(\det A_m^2)^{p-1}$, so that, if $(*)$ holds for m, we have

$$\det A_c^2 = p^{(p-2)\phi(m)}\left\{(-1)^{\phi(m)/2}\prod_{q|m} q^{(q-2)\phi(m/q)}\right\}^{p-1}$$
$$= p^{(p-2)\phi(c/p)}\prod_{q|m} q^{(q-2)\phi(c/q)},$$

where q runs over the prime factors of m. Thus our induction is complete in the case $c = pm$, p odd, $p|m$, $m > 2$, and so $(*)$ is proved.

Next, assume that $c = p^n$, p prime, $n > 1$. Then

$$C_c(j+k) = \begin{cases} 0 & \text{if } (c, j+k) < p^{n-1} \\ -p^{n-1} & \text{if } (c, j+k) = p^{n-1} \\ p^n - p^{n-1} & \text{if } (c, j+k) = p^n. \end{cases}$$

In this case the sum of the elements in each row of A_c^2 is zero and so $\det A_c^2 = 0$.

Finally, assume that $c = p^n m$, p prime, $n > 1$, $p \nmid m$. Since

$$C_c(j+k) = C_{p^n}(j+k) \cdot C_m(j+k),$$

we see again that the sum of the elements in each row of A_c^2 is zero and so $\det A_c^2 = 0$.

REFERENCES

1. Tom Apostol, Introduction to Analytic Number Theory, Springer-Verlag, 1976.
2. Oystein Ore, Number Theory and Its History, McGraw-Hill, 1948.

Solved also by the proposer.

Homeomorphisms of Compact Metric Spaces

6612 [1989, 846]. *Proposed by Ebrahim Salehi, University of Nevada, Las Vegas.* Suppose X is a compact metric space with metric d and suppose $T: X \to X$ is continuous. If

$$\inf_{n \in N} d(T^n x, T^n y) > 0$$

for each pair x, y of distinct elements of X, prove that T is onto.

Solution by Eugene Curtin, Southwest Texas State University, San Marcos, TX. We establish that T is a homeomorphism of X onto X without using the assumption that T is continuous. Let T^0 denote the identity map on X and define $d': X \times X \to \mathbb{R}$ by

$$d'(x, y) = \inf_{N \cup \{0\}} d(T^n x, T^n y).$$

The hypothesis on T shows that $d'(x, y) = 0$ if and only if $x = y$. The other metric space axioms are trivial to check on (X, d'). For each x, y in X we have $d(x, y) \geqslant d'(x, y)$, so the identity map $T^0: (X, d) \to (X, d')$ is continuous. Since a continuous bijection $f: X \to Y$ of a compact metric space X onto a metric space Y is a homeomorphism (see [1, p. 98]), (X, d') is topologically equivalent to (X, d); in particular, it is compact. Note that $d'(Tx, Ty) \geqslant d'(x, y)$ for all x, y in X. Hence it

REFERENCE

1. E. T. Copson, Metric Spaces, Cambridge University Press, Cambridge, 1968.

Solved also by J. M. Borwein & A. C. Thompson (Canada), Nathan S. Feldman, Michael Fichter (Germany), Ilias Kastanas, Daniel J. Opitz, Harold Steinmacker (Austria) and the proposer. One incomplete solution also was received.

A Problem Related to Engel's 1 + ab Problem

6619 [1990, 75]. *Proposed by Arthur L. Holshouser, Charlotte, NC, and Benjamin G. Klein, Davidson College, NC.*

Suppose that $N = 4^h(8k + 7)$ for non-negative integers h and k.

(a) Show that there exist infinitely many distinct 5-tuples (a, b, c, d, e) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2 + e^2)/(abcde + 1). \qquad (*)$$

(b) Show that there is no 4-tuple (a, b, c, d) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2)/(abcd + 1). \qquad (**)$$

Solution by R. J. Chapman, University of Exeter, U.K. It is well known that such integers N are exactly the nonnegative integers not expressible as sums of three integral squares, and that every nonnegative integer is the sum of four integral squares (Lagrange's theorem). We shall also use the trivial fact that each such N is at least 7.

For (a), suppose that (a, b, c, d, e) is a solution of the given Diophantine equation $(*)$ with $0 \leqslant a \leqslant b \leqslant c \leqslant d \leqslant e$ and $b > 0$. We claim there is an integer $a' > a$ such that (a', b, c, d, e) is also a solution. Clearly

$$a^2 - (Nbcde)a + (b^2 + c^2 + d^2 + e^2 - N) = 0$$

is a monic quadratic equation in a whose other root

$$a' = Nbcde - a$$

is an integer such that (a', b, c, d, e) is also a solution to $(*)$. If $a' \leqslant a$, then

$$2a \geqslant Nbcde \geqslant 7b > 2b \geqslant 2a,$$

a contradiction, so $a' > a$. Now by Lagrange's theorem

$$N = (0^2 + b_0^2 + c_0^2 + d_0^2 + e_0^2)/(1 + 0 \cdot b_0c_0d_0e_0),$$

where $0 < b_0 \leqslant c_0 \leqslant d_0 \leqslant e_0$. Hence our procedure has a starting point. It next yields a solution $(a_0, b_0, c_0, d_0, e_0)$ in positive integers. By rearranging and repeating the procedure we obtain an infinite family of solutions, since $a + b + c + d + e$ increases at each stage.

For (b) let (a, b, c, d) be a solution of the given Diophantine equation $(**)$ in positive integers with $a \leqslant b \leqslant c \leqslant d$. We claim there exists an integer d' with $0 \leqslant d' < d$ such that (a, b, c, d') is also a solution. This would yield a contradiction, since by rearranging and repeating we must eventually get $d' = 0$, i.e., that N is the sum of three squares. Let $d' = Nabc - d$. By an argument similar to that of (a) we find that (a, b, c, d') also solves $(**)$. Since

$$N(1 + abcd') = a^2 + b^2 + c^2 + d'^2,$$

REFERENCE

1. E. T. Copson, Metric Spaces, Cambridge University Press, Cambridge, 1968.

Solved also by J. M. Borwein & A. C. Thompson (Canada), Nathan S. Feldman, Michael Fichter (Germany), Ilias Kastanas, Daniel J. Opitz, Harold Steinmacker (Austria) and the proposer. One incomplete solution also was received.

A Problem Related to Engel's 1 + ab Problem

6619 [1990, 75]. *Proposed by Arthur L. Holshouser, Charlotte, NC, and Benjamin G. Klein, Davidson College, NC.*

Suppose that $N = 4^h(8k + 7)$ for non-negative integers h and k.

(a) Show that there exist infinitely many distinct 5-tuples (a, b, c, d, e) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2 + e^2)/(abcde + 1). \qquad (*)$$

(b) Show that there is no 4-tuple (a, b, c, d) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2)/(abcd + 1). \qquad (**)$$

Solution by R. J. Chapman, University of Exeter, U.K. It is well known that such integers N are exactly the nonnegative integers not expressible as sums of three integral squares, and that every nonnegative integer is the sum of four integral squares (Lagrange's theorem). We shall also use the trivial fact that each such N is at least 7.

For (a), suppose that (a, b, c, d, e) is a solution of the given Diophantine equation $(*)$ with $0 \leqslant a \leqslant b \leqslant c \leqslant d \leqslant e$ and $b > 0$. We claim there is an integer $a' > a$ such that (a', b, c, d, e) is also a solution. Clearly

$$a^2 - (Nbcde)a + (b^2 + c^2 + d^2 + e^2 - N) = 0$$

is a monic quadratic equation in a whose other root

$$a' = Nbcde - a$$

is an integer such that (a', b, c, d, e) is also a solution to $(*)$. If $a' \leqslant a$, then

$$2a \geqslant Nbcde \geqslant 7b > 2b \geqslant 2a,$$

a contradiction, so $a' > a$. Now by Lagrange's theorem

$$N = \left(0^2 + b_0^2 + c_0^2 + d_0^2 + e_0^2\right)/(1 + 0 \cdot b_0c_0d_0e_0),$$

where $0 < b_0 \leqslant c_0 \leqslant d_0 \leqslant e_0$. Hence our procedure has a starting point. It next yields a solution $(a_0, b_0, c_0, d_0, e_0)$ in positive integers. By rearranging and repeating the procedure we obtain an infinite family of solutions, since $a + b + c + d + e$ increases at each stage.

For (b) let (a, b, c, d) be a solution of the given Diophantine equation $(**)$ in positive integers with $a \leqslant b \leqslant c \leqslant d$. We claim there exists an integer d' with $0 \leqslant d' < d$ such that (a, b, c, d') is also a solution. This would yield a contradiction, since by rearranging and repeating we must eventually get $d' = 0$, i.e., that N is the sum of three squares. Let $d' = Nabc - d$. By an argument similar to that of (a) we find that (a, b, c, d') also solves $(**)$. Since

$$N(1 + abcd') = a^2 + b^2 + c^2 + d'^2,$$

d' must be nonnegative. If $d' \geqslant d$, then $2d \leqslant Nabc$ or

$$2d(1 + abcd) \leqslant abc(a^2 + b^2 + c^2 + d^2),$$

so that

$$0 < 2d \leqslant abc(a^2 + b^2 + c^2 - d^2).$$

From $a \leqslant b \leqslant c$ we deduce $d^2 < 3c^2$ and

$$N + Nabc^2 \leqslant N(1 + abcd) = a^2 + b^2 + c^2 + d^2 < 6c^2.$$

But $N \geqslant 7$, so this is false, Hence $d' < d$ and the result follows.

Editorial comment. Both R. J. Chapman and Christos Athanasiadis remarked on the similarity of this problem to the celebrated "$1 + ab$ problem" of Arthur Engel (from the 29th International Mathematical Olympiad). Chapman remarks that the above method of proof yields the following result. If $a_1, \ldots, a_r$ are non-negative integers and

$$N = \frac{a_1^2 + \cdots + a_r^2}{1 + a_1 a_2 \cdots a_r}$$

is an integer, then N is the sum of $(r - 1)$ squares of integers. Engel's problem is the case $r = 2$ and part (b) of the given problem is the case $r = 4$.

Solved also by Christos Athanasiadis (student) and the proposers.

Subsets of the Plane Intersecting Every Straight Line in Exactly Two Points

6624 [1990, 162]. *Proposed by K. A. Ross, University of Oregon, Eugene.*

(i) Show that there is a subset A of the plane R^2 that intersects every straight line in exactly two points.

(ii) Can the set A be selected to be Lebesgue measurable?

(iii)* Can the set A be selected to be a Borel set?

Partial solution by Robert High, New York City and Fred Galvin, University of Kansas. For (i) let κ be the least ordinal of cardinality equal to the continuum (we do not require the *continuum hypothesis*). Let the set of all lines in the plane be placed in a 1–1 correspondence with the set of ordinals less than κ. Now let S begin as the empty set. For each line l_λ with $\lambda < \kappa$ we will adjoin at most two points to S so that at each stage the cardinality of S (namely, the cardinality of λ) will be strictly less than κ. Consider all lines determined by points already in S. Their number is at most the square of the cardinality of λ, and this is strictly less than the cardinality of κ. If l_λ is among the lines determined by points already in S, then no points are added to S for l_λ. Otherwise, the set G_λ of points on l_λ but not on any line determined by points already in S is infinite (as the family of such lines has cardinality strictly less than κ, and each meets l_λ in at most one point). Adjoin one or two points from G_λ to S, as required to ensure that S has exactly two points of l_λ. Note, by definition of G_λ, that this action does not add to S any points on any previously considered line l_β with $\beta < \lambda$. Analogously, no future action will add points on l_λ to S. Continue this process transfinitely: the resulting S is the desired set A.

For (ii), use the same construction restricted to a measure zero subset of the plane that meets every line in continuum many points. For example, use $(R \times C) \cup (C \times R)$ where C is the usual Cantor set of measure zero. Vertical lines intersect $R \times C$ in continuum many points. All the other lines intersect $C \times R$ in continuum many points. Construct A as a subset of this set; this ensures that it, too, has Lebesgue measure zero.

Editorial comment. In his paper "Infinite combinatorics and definability," *Annals of Pure and Applied Logic* 41 (1989), 179–203, Arnold W. Miller points out (p. 193) that (i) is a theorem of S. Mazurkiewicz, *C. R. Sc. et Lettres de Varsovie* 7 (1914), 322–383, that (iii) is a recognized open problem, but that in D. G. Larman, "A problem of incidence," *J. London Math. Soc.* 43 (1968), 407–409, it is shown that no "two-point set" can be an F_σ set. Miller (p. 195) also mentions the unpublished result of D. Mauldin that two-point sets are totally disconnected. The editors thank Fred Galvin for bringing this reference to their attention.

Solved also by the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Beyond the Third Dimension. By Thomas F. Banchoff, Scientific American Library, New York, 1990, xii + 210 pp.

H. S. M. COXETER
Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

This beautiful book gently introduces the reader to the idea of many dimensions. Its abundant pictures include portraits of Albert Einstein and Marston Morse, Salvador Dali and the author, Immanuel Kant, and K. F. Gauss. On page 2 there is a facsimile of the title page of the first edition of *Flatland* [1], "Price half-a-crown," which includes

O day and night, but this is wondrous strange
And therefore as a stranger give it welcome.

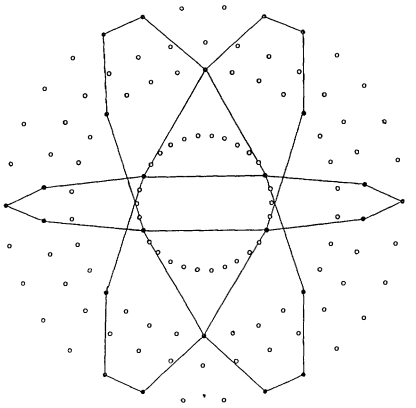
Many of the illustrations are elegantly coloured. Those on page 89 are reminiscent of Oliver Byrne [2] which was dedicated "to the right honourable the Earl of Fitzwilliam, etc. etc. etc." In contrast, the author reminds the reader (on page 102) that Schläfli [6] published "a long work on higher-dimensional geometry which contained not one picture." Such lack of human consideration was repeated by von Staudt [7; 8].

Page 78 includes a photograph showing the distribution of orifices on a pollen grain, which explains why the problem of packing disks (or "caps") on a sphere is commonly ascribed to P. M. L. Tammes, a botanist.

Page 86 introduces the "Greek Geometry Game" by a coloured drawing of Max Bill's polygonal spiral, which consists of $n - 1$ sides, all having the same length, of each regular n-gon for $n = 3, 4, 5, 6, 7, 8$. The Greeks used their favourite instruments, the ruler and compasses, to construct all these polygons except the heptagon. The impossibility of so constructing the regular heptagon was finally established by Wantzel [9].

Among the most striking objects "beyond the third dimension" are the 4-dimensional analogues of the five Platonic solids. These are the six convex regular polytopes, which the author begins to describe in Chapter 5, paying special attention to the 600-cell, whose six hundred facets or cells are regular tetrahedra [3, pp. 247–281; 4, p. 42]. The figure on page 100 is a projection into a plane, of 600-cell's 120 vertices, 360 of its 720 edges, 120 of its 1200 faces, and 36 of its 72 "equatorial" decagons (analogous to the three "equatorial" squares in the regular octahedron). Those 120 faces appear undistorted as *equilateral* triangles, of various sizes because of different foreshortening [4, p. 48]. The accompanying figure shows two antipodal triangles and the three decagons that contain their sides.

The decagons appear in two shapes, typical of 15 + 15 of the 36. The remaining 6 appear as undistorted regular decagons. In fact, the author's figure can be derived by copying the reviewer's figure and rotating it about the center through all multiples of 12°. (In this process, some edges will be repeated.)



The 600-cell's remaining 360 edges, which the author omitted, form another set of 36 equatorial decagons, quite different because they form no equilateral triangles. He returns to the subject of regular polytopes in Chapter 6, where he reproduces "perspective" views of the 5-cell, 16-cell and 24-cell from Hilbert and Cohn-Vossen [5, pp. 150–152], pointing out that these views are analogous to Schlegel diagrams.

Being an expert on computer graphics, the author makes striking use of that technique in showing how solids and hypersolids change their apparent shape when they are rotated.

This colorful book, with its three pages of "Further readings" and useful index, will be appreciated by the thousands of people who are curious about the concept of hyperspaces, and by the still greater number of people who enjoy beautiful pictures.

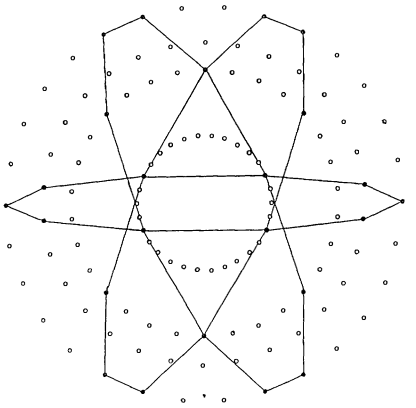
REFERENCES

1. E. A. Abbott, Flatland, A Romance of Many Dimensions By A Square, Seeley, London, 1884.
2. Oliver Byrne, The first six books of the Elements of Euclid, in which coloured diagrams and symbols are used instead of letters for the greater ease of learners, W. Pickering, London, 1847.
3. H. S. M. Coxeter, Regular Polytopes (3rd. ed.), Dover, New York, 1973.
4. _____, Regular Complex Polytopes (2nd. ed.), Cambridge University Press, 1991.
5. David Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.
6. Ludwig Schläfli, Theorie der vielfachen Kontinuität, in *Gesammelte Mathematische Abhandlungen*, I, Birkhäuser, Basel, 1950.
7. G. K. C. von Staudt, Geometrie der Lage, Nürnberg, 1847.
8. _____, Beiträge zur Geometrie der Lage, Nürnberg, 1857.
9. L. Wantzel, Recherches sur les moyens de reconnaître si un Probème de Géométrie peut se résoudre avec la règle et le compas, *J. de Math. pures et appl.*, 2 (1837) 366–372.

Global Differential Geometry. Editor, S. S. Chern. MAA Studies in Mathematics, Volume 27, The Mathematical Association of America, Washington, DC, 1989, ii + 254 pp.

THOMAS F. BANCHOFF
Department of Mathematics, Brown University, Providence, RI 02912

This book is an excellent collection of articles for teachers and students who are looking for a supplement to a standard geometry undergraduate curriculum, or



The 600-cell's remaining 360 edges, which the author omitted, form another set of 36 equatorial decagons, quite different because they form no equilateral triangles. He returns to the subject of regular polytopes in Chapter 6, where he reproduces "perspective" views of the 5-cell, 16-cell and 24-cell from Hilbert and Cohn-Vossen [5, pp. 150–152], pointing out that these views are analogous to Schlegel diagrams.

Being an expert on computer graphics, the author makes striking use of that technique in showing how solids and hypersolids change their apparent shape when they are rotated.

This colorful book, with its three pages of "Further readings" and useful index, will be appreciated by the thousands of people who are curious about the concept of hyperspaces, and by the still greater number of people who enjoy beautiful pictures.

REFERENCES

1. E. A. Abbott, Flatland, A Romance of Many Dimensions By A Square, Seeley, London, 1884.
2. Oliver Byrne, The first six books of the Elements of Euclid, in which coloured diagrams and symbols are used instead of letters for the greater ease of learners, W. Pickering, London, 1847.
3. H. S. M. Coxeter, Regular Polytopes (3rd. ed.), Dover, New York, 1973.
4. _____, Regular Complex Polytopes (2nd. ed.), Cambridge University Press, 1991.
5. David Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.
6. Ludwig Schläfli, Theorie der vielfachen Kontinuität, in *Gesammelte Mathematische Abhandlungen*, I, Birkhäuser, Basel, 1950.
7. G. K. C. von Staudt, Geometrie der Lage, Nürnberg, 1847.
8. _____, Beiträge zur Geometrie der Lage, Nürnberg, 1857.
9. L. Wantzel, Recherches sur les moyens de reconnaître si un Probème de Géométrie peut se résoudre avec la règle et le compas, *J. de Math. pures et appl.*, 2 (1837) 366–372.

Global Differential Geometry. Editor, S. S. Chern. MAA Studies in Mathematics, Volume 27, The Mathematical Association of America, Washington, DC, 1989, ii + 254 pp.

THOMAS F. BANCHOFF
Department of Mathematics, Brown University, Providence, RI 02912

This book is an excellent collection of articles for teachers and students who are looking for a supplement to a standard geometry undergraduate curriculum, or

who want to get an idea about the key problems that are informing this important subject. Twenty years ago, the first edition of this work appeared, including earlier versions of many of the same articles. The best way to appreciate this book is to read the introduction to the first edition, which was inexplicably left out of the second. Many books do print both versions, with the first giving an overview of the original project and the second keying into the new developments in the field, and that approach would clear up the question about why certain articles have been included. The original title of the MAA Studies volume was *Studies in Global Geometry and Analysis* and the first entry was Marston Morse's short expository article "What Is Analysis in the Large?" This was followed by Shiing-Shen Chern's Chauvenet prize-winning essay "Curves and Surfaces in Euclidean Space," and together these articles introduced the two notions of analysis and classical differential geometry which came together to form global differential geometry. The new title *Global Differential Geometry* reflects a shift in emphasis and Morse's essay does not serve any longer as the introduction to the subject. Without some explanation in the introduction, it seems totally out of place near the end of this new volume.

Beginning readers should probably read the articles in this book in a different order from their sequence in the table of contents. The best beginning is still the article by Chern on curves and surfaces. It can then be followed by the article by Harley Flanders on Differential Forms, which was very good in the first edition and it is even better now, representing a short version of his soon-to-be-reprinted book on the same subject. Only then should the reader proceed to the article which is now the lead, "Vector Bundles with a Connection." Without a good understanding of the classical differential geometry of curves and surfaces, a reader will have difficulty following the first article as it moves rather quickly into a fairly abstract framework. Within a few pages, the reader is confronted with "the vector space of sections of the tensor product of a vector bundle with the cotangent space of a manifold." The notions of "tensor product" and "cotangent space" are not defined in the article. Readers not familiar with these concepts from a course in algebra or advanced calculus are likely to be lost at this point. Of course it is impossible to include all relevant background for a vast subject in a short survey article, but terminology can stand in the way of the author's stated aim of showing that "a connection on a vector bundle is basically a very elementary concept."

The original volume was somewhat uneven, and that problem remains in the new one. The article of Lamberto Cesari on "Surface Area" is still quite technical (due to the technical nature of the subject) and the article by L. A. Santaló on "Integral Geometry" is still mysterious (due to the mysterious nature of the subject). The text of neither article has been changed in the last twenty years, although some recent references have been appended in each case. The same is true for the article by Shoshichi Kobayashi "On Conjugate and Cut Loci."

Three new articles are especially welcome, presenting an exposition of recent work in several different areas: Robert Osserman's essay "Minimal Surfaces in R^3," Hermann Karcher's article "Riemannian Comparison Theorems," and Patrick Eberlein's "Manifolds of Nonpositive Curvature." The first of these uses only notions of advanced calculus and a first course in complex variables, so it should be accessible to advanced undergraduates. The second and third bring in numer-

ous concepts from modern differential geometry, and both make an attempt to be self-contained, so they represent a good introduction into the formalism as well as the concepts of global differential geometry.

All in all, this volume, like its predecessor, serves a very important purpose, that of introducing a broad range of readers to the methods and scope of global differential geometry, a field that will continue to grow in significance. What, we may wonder, will the next edition of this book look like twenty years from now?

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, L. *CRC Handbook of Mathematical Curves and Surfaces.* David H. von Seggern. CRC Pr, 1990, 286 pp, $49.95. [ISBN: 0-8493-0155-6] Pictures of families of functions on right-hand pages, facing displays of formulas with relevant parameters: algebraic curves, transcendental functions, orthogonal and non-orthogonal polynomials, special functions (e.g., Legendre, Bessel), probability densities, catastrophe curves, spirals and helices, algebraic and transcendental surfaces, and nondifferentiable curves and surfaces (including polygons and polyhedra). A handy reference for those without a computer. LAS

General, S*(15-18), P*, L*. *Problems in Applied Mathematics: Selections from SIAM Review.* Ed: Murray S. Klamkin. SIAM, 1990, xxv + 588 pp, $36.50 (P). [ISBN: 0-89871-259-9] A compilation of 380 of *SIAM Review's* most interesting problems dating back to the journal's inception in 1959. This outstanding collection of instructive problems is organized into twenty-two broad sections such as probability, combinatorics, series, special functions, and so forth. Most of the problems include motivational material describing how the problem arose; the solutions, some of which approach the status of short papers, are accessible to a wide audience (say, at the level of the *American Mathematical Monthly*). Solutions generally include references to further literature and pertinent comments by Klamkin. There is something in this volume for everyone, and is a welcome addition to the problems literature. LCL

General, S(13-14), L.** *Conquering Mathematics: From Arithmetic to Calculus.* Lloyd Motz, Jefferson Hane Weaver. Plenum Pr, 1991, xiv + 305 pp, $23.50. [ISBN: 0-306-43768-6] An informal discursive survey of mathematical basics—what many people should have learned in school but didn't—beginning with a thorough discussion of the whys and wherefores of arithmetic. Focuses throughout, and especially at the end in an epilogue on the relations between mathematics and physics, on the "predictive power" of mathematics. Written for literate readers who have forgotten (or never learned) basic mathematics but can nevertheless comprehend logically and linguistically sophisticated exposition. LAS

General, S, L*. *Beyond Numeracy: Ruminations of a Numbers Man.* John Allen Paulos. Alfred A. Knopf, 1991, xiii + 285 pp, $22. [ISBN: 0-394-58640-9] Cultural literacy for mathematics: seventy brief topical essays, arranged alphabetically, covering mathematical vocabulary from algebra to Zeno including expected topics such as binary numbers, chaos, differential equations, prime numbers, statistics, and topology as well as unexpected diversions (ethics, human consciousness, humor, even "Oulipo"). Ends with a chronological list of the "top forty" mathematicians from antiquity to the present, each with brief identifiers, and an afterword reflection on the fact that only White males made the list. LAS

General, S? *Only Problems, Not Solutions!* Florentin Smarandache (POB 42561, Phoenix, AZ 85080), 1991, 37 pp, (P). [ISBN: 1-879585-00-6] Thirty-five easily stated, but probably difficult, unsolved problems involving integers, sequences, equations, configurations. Many are generalizations of known results (references included). Virtually no discussion. LCL

General, P. *Winning Women Into Mathematics.* Eds: Patricia Clark Kenschaft, Sandra Keith. MAA, 1991, ix + 78 pp, $11 (P). [ISBN: 0-88385-453-8] A very varied collection of materials gathered by the MAA Committee on the Participation of Women that documents both the increasing role of women in mathematics as well as the continuing impediments they face. Chronicles committee activities (including "micro-inequity" skits they have arranged at meetings); highlights (with many photographs) numerous "firsts" for women in mathematics; lists cultural impediments that women face in mathematical careers (e.g., "women may also be more reluctant to write in a style that seems to embody conceit"); suggests actions; and describes numerous intervention projects. LAS

General, S, C**, P*.** *Theorist.* (Version 1.11). Allan Bonadio. Prescience (939 Howard St., No. 306, San Francisco, CA 94103), 1991, $399.95. *Learning Guide,* iv + 242 pp, (P); *Reference Manual,* xiii +

360 pp, (P). A powerful yet intuitive Macintosh package for symbolic manipulation and visualization that provides a handy electronic blackboard for anyone using mathematics. Symbol palette eases equation entry. Menu options control manipulations (factor, substitute, transform simplify, expand, etc.), while window buttons control graphing (zooming, resolution, resizing, perspective, algebraic details). Animation option creates and saves frames in PICS format and then redisplays them in rapid sequence. Employs notebooks to store equations, graphs, and manipulations; comments containing text or graphics can be added with button-controlled display. Results can be saved, edited, and printed in standard Macintosh style. Greater ease, less power than larger symbolic algebra systems; excellent documentation in two parts—a tutorial guide to graphing, algebra, and calculus, and a reference manual with careful explanations of all options. LAS

General, S*, P, L.** *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* Manfred Schroeder. WH Freeman, 1990, xviii + 429 pp, $32.95. [ISBN: 0-7167-2136-8] An inviting exposition of the worldly and mathematical manifestations of self-similarity: music and "pink noise;" roulette and pinball machines; quasiperiodic matter and renormalization; and, of course, fractals in all their well-known forms. Illustrates with numerous examples the fascinating interplay between the mathematical and physical consequences of changes in scale and self-similarity. Clearly written, using formulas and graphs effectively for a literate but not highly scientific audience. LAS

General, P. *Mathematical Sciences, Technology, and Economic Competitiveness.* Ed: James G. Glimm. National Academy Pr, 1991, x + 114 pp, (P). [ISBN: 0-309-04483-9] An important innovative study of the role of mathematics, especially of simulation and modelling, in U.S. industries, from economic planning through product design, manufacturing, and quality control to inventories, marketing, and maintenance. Examples drawn from diverse industries illustrate much useful mathematics: statistical quality improvement, discrete and combinatorial optimization, simulation, and probabilistic models. LAS

General, S, P. *Environmental Risk Reporting: The Science and the Coverage.* Eds: Sharon M. Friedman, Carol L. Rogers. Lehigh Univ (Dept. of Journalism, Univ. Center 29, Bethlehem, PA 18015), 1991, 48 pp, (P). Brief papers from a day-long AAAS symposium on the journalistic challenges of covering environmental risks: how to explain and quantify risks in a way that is meaningful and helpful to the public. Little explicit mathematics, but much valuable insight into how ordinary people (journalists and their audiences) (mis)interpret simple statements of magnitude, rates, and probabilities. LAS

General, S*, L*. *Flatland: A Romance of Many Dimensions.* Edwin Abbott Abbott. Princeton Univ Pr, 1991, 103 pp, $7.95 (P). [ISBN: 0-691-02525-8] A new edition of a classic social satire, at least the dozenth that has appeared since the original was published in 1884. Tom Banchoff's introduction reveals diverse issues from Abbott's agenda that are still of interest today ranging from parochial perspectives (e.g., treatment of women) to visualization through slicing (e.g., exploratory data analysis, CAT scanners). A timely republication well-suited to today's computer-graphics generation. LAS

Elementary, S, L. *The Method of Coordinates.* I.M. Gel'fand, E.G. Glagoleva, A.A. Kirillov. Birkhauser Boston, 1990, x + 73 pp, $14.50 (P). [ISBN: 0-8176-3533-5] Second volume in the newly re-published series of booklets from Moscow State University's correspondence school for rural secondary school pupils. Introduces Cartesian coordinates in 2, 3, and 4 dimensions, with tangential discussions of other systems. First English edition published in 1967 by MIT Press (TR, October 1967). LAS

Mathematics Appreciation, T(13: 1), S. *Mathematics and Global Survival, Second Edition.* Richard H. Schwartz. Ginn Pr, 1990, xvi + 197 pp, (P). [ISBN: 0-536-57540-1] Reviews a wide variety of elementary mathematical topics (e.g., percents, ratios and proportions, graphical methods, arithmetic and geometric sequences, probability, and statistics) using "real-world" data on population growth, the arms race, energy utilization, waste management, and other issues. Includes a bibliography, a list of possible term paper topics, and solutions to many of the exercises. AO

Mathematics Appreciation, T(13-14: 1), S. *Universal Patterns, Book One, Revised Edition.* Martha Boles, Rochelle Newman. Pythagorean Pr, 1990, xxi + 266 pp, $29.95 (P). [ISBN: 0-9614504-3-6] Main thrust is to show relationships among things which appear disconnected. Focuses on nature, mathematics, and art. Mathematics is used to define abstract relationships found in nature while art can create relationships in nature. Both art and mathematics can interpret nature and nature can suggest directions in both mathematics and art. A treasure of geometric constructions fills the text with accompanying discussion of related natural and artistic phenomena. Chapters: basic constructions, unique relationships, special triangles, dynamic rectangles, Fibonacci numbers, spirals. Lots of problems and suggested projects. Good references; appendices; answers to selected problems; glossary; index. RJA

Mathematics Appreciation, S. *Lecture Notes in Mathematics-1458: Global Solution Branches of Two Point Boundary Value Problems.* Renate Schaaf. Springer-Verlag, 1990, xix + 141 pp, $20 (P). [ISBN: 0-387-53514-4] Introduces the parameterized equations for the two point boundary value problems, and the time map T of the equations; derives a time map formula for branches bifurcation which overcomes the singularity; covers applications and results, including Newmann problems and Dirichlet's problem; also

studies the stability of the time map and some computer plots. BL

Finite Mathematics, T(13). *Finite Mathematics.* William B. Owen, Wm. Frederick Cutlip. Harcourt Brace Jovanovich, 1991, xlix + 678 pp, $35 net. [ISBN: 0-15-527546-1] For an applications-oriented course for students of business, the life sciences, and the social sciences, assuming one year of high school algebra. Covers linear mathematics, probability, statistics, and mathematics of finance with MINITAB examples. JPH

Education, P, L*. *What Works: Building Natural Science Communities.* Project Kaleidoscope (ICO, 1730 Rhode Island Ave., NW, Suite 1205, Washington, DC 20036), 1991, xviii + 100 pp, (P). An analysis of productive undergraduate programs in science and mathematics, stressing mechanisms that make learning experiential, personal, and rooted in "natural science communities." Includes chapters on science, mathematics, human resources, faculty, and facilities; texts of colloquium presentations by national policy leaders; and vignettes of illustrative programs. LAS

Education, P, L*. *Math Panic.* Laurie Buxton. Heinemann, 1991, xv + 260 pp, (P). [ISBN: 0-435-08313-9] A startling account of anxiety and despair recalled by British adults as they reflected on their struggles with mathematics, both in school and at work. Author Buxton builds from these experiences a very believable analysis of how reason is frequently blocked by emotion, leading to practical suggestions for mathematics teachers at all levels. Revised American edition of a book originally published in England in 1981. LAS

Education, P, L. *The State of Mathematics Achievement.* Ina V.S. Mullis, *et al.* Educational Testing Service, 1991, 532 pp, (P). [ISBN: 0-88685-15-7] Detailed report of the 1990 national assessment of mathematics at grades 4, 8, and 12, including trial state-wide data at eighth grade. Findings generally confirm earlier studies: only 5% of high school seniors understand geometry and algebra; fewer than half have consistent command of decimals, percents, and fractions. At the eighth grade level, states with fewer urban areas, fewer disadvantaged students and fewer minority students performed above average, as did students who watched less television, read more at home, and attended school regularly. Policy factors such as dollars per student, required courses, teachers' salaries, and length of school year appear to be uncorrelated with performance. LAS

Education, P*, L. *Mathematical Modeling in the Secondary School Curriculum: A Resource Guide of Classroom Exercises.* Eds: Frank Swetz, J.S. Hartzler. NCTM, 1991, viii + 136 pp, $9.50 (P). [ISBN: 0-87353-306-2] Presents twenty-two successful classroom modeling exercises developed and class-tested under the auspices of the Mathematical Sciences Program of the Pennsylvania State University at Harrisburg. The mathematics required ranges from elementary through precalculus. Also includes teaching hints and additional information (e.g., a bibliography) for secondary school instructors interested in teaching mathematical modeling. AO

History, S*, P, L*. *Ethnomathematics: A Multicultural View of Mathematical Ideas.* Marcia Ascher. Brooks/Cole, 1991, ix + 203 pp, $38.95. [ISBN: 0-534-14880-8] An innovative study of mathematical ideas (e.g., numbers, space, logic, patterns, chance) embedded in the cultural traditions of "small scale" societies (Carolina Islands, Navajo, Maori, Warlpiri, ...). Intended to demonstrate that "mathematics" is a distinctly Western construct, but that "mathematical ideas" are universal and culturally dependent. Thoroughly researched and well-documented; concluding chapter on implications treads on controversial issues of values, education, and purpose. LAS

History, P*, L.** *Augustin-Louis Cauchy: A Biography.* Bruno Belhoste. Transl: Frank Ragland. Springer-Verlag, 1991, xii + 380 pp, $79. [ISBN: 0-387-97220-X] A fascinating account of the professional career of Cauchy, an arrogant, argumentative, rigid, pro-cleric royalist whose immense mathematical contributions occurred in the midst of successive revolutions and counter-revolutions in nineteenth-century France. Cauchy's political views and personality led to his repeated rejection for academic positions; his reputation as a terrible teacher caused repeated conflicts with university authorities; yet his drive for rigorous truth made him the greatest French mathematician of his generation. Carefully researched with extensive notes pointing to original manuscripts, this biography offers an effective balance of historical and mathematical detail, including a valuable chapter that explores the philosophical rationale of Cauchy's uncompromising approach to politics, religion, and mathematics. LAS

History, L*. *Rabdology.* John Napier. Transl: William Frank Richardson. Charles Babbage Institute Reprint Ser. for History of Computing, V. 15. MIT Pr, 1990, xxxvii + 135 pp, $40. [ISBN: 0-262-14046-2] The original user manual: the first complete English translation of Napier's 1617 Latin *Rabdologia,* a guide to the use of "Napier's Bones," a "high-speed promptuary for multiplication." From the Dedicatory Epistle: "To perform calculations is a difficult and lengthy process, the tedium of which deters many from the study of mathematics." In addition to lengthy but very clear instructions for multiplication, division, and extraction of roots, Napier's brief treatise contains the first written reference to the decimal point, and a remarkable geometric algorithm for "location arithmetic on a chessboard" based on binary numbers developed "as more of a lark than a labor." LAS

History, T(14-17), S, L. *A History of Mathematics, Second Edition.* Carl B. Boyer, Uta C. Merzbach. Wiley, 1991, xx + 715 pp, $24.95 (P). [ISBN: 0-471-54397-7] A text essentially unchanged from the 1989 revised edition (TR, October 1989). That edition, however, was designed for classroom use. In this edition the exercises have been dropped

with the 'outside-the-lecture-room' reader in mind. (1968 edition, TR, January 1969; 1985 Princeton Univ. Pr. edition, TR, February 1986.) KS

History, S(13-18), P, L. *The Development of Newtonian Calculus in Britain, 1700-1800.* Niccolò Guicciardini. Cambridge Univ Pr, 1989, xii + 228 pp. [ISBN: 0-521-36466-3] Devoted to the "calculus of fluxions," this book argues that Britain did not experience the post-Newton mathematical decline that it is reputed to have suffered. Examines research and teaching of this calculus during 18th century; includes sociological discussions. Considers importance of work of, among others, Newton, Cotes, Taylor, Stirling, Maclaurin, Simpson. KS

History, L.** *Women in Science: Antiquity through the Nineteenth Century: A Biographical Dictionary with Annotated Bibliography.* Marilyn Bailey Ogilvie. MIT Pr, 1988, xiii + 254 pp, $12.50 (P). [ISBN: 0-262-15031-X] The introduction is a historical essay about science and women. Each biographical sketch includes lists of major work by the subject herself and works about her. A good source book for courses on women and science. MLR

Logic, P. *Classification Theory and the Number of Non-Isomorphic Models, Second Revised Edition.* S. Shelah. Stud. in Logic & Found. of Math., V. 92. North-Holland (US Distr: Elsevier Science), 1990, xxxiv + 705 pp, $242.75. [ISBN: 0-444-70260-1] This edition includes corrections, list of open problems, and four new chapters covering developments in classification theory since the *First Edition* (TR, August-September 1979). Note price! KES

Graph Theory, P. *Random Graphs '87.* Eds: Michał Karoński, Jerzy Jaworski, Andrzej Ruciński. Wiley, 1990, ix + 368 pp, $84.95. [ISBN: 0-471-92749-X] Twenty-four expository and research papers from the Third International Seminar on Random Graphs and Probabilistic Methods in Combinatorics held at Poznań, 1987, plus open problems. JPH

Graph Theory, P. *Topics on Domination.* S.T. Hedetniemi, R.C. Laskar. Annals of Disc. Math., V. 48. North-Holland (US Distr: Elsevier Science), 1991, 280 pp, $143. [ISBN: 0-444-89006-8] A comprehensive collection on the theory of dominating sets in graphs. Twenty-one papers on the standard domination number, on new variations, and on algorithmic topics, plus a bibliography. JPH

Combinatorics, S(18), P. *Submodular Functions and Optimization.* Satoru Fujishige. Annals of Disc. Math., V. 47. North-Holland (US Distr: Elsevier Science), 1991, ix + 270 pp, $97.25. [ISBN: 0-444-88556-0] Exposition of submodular function theory and its constructive aspects. Treats submodular systems, base polyhedra, network flow problems, the intersection problem, connections between submodular functions and convex analysis, nonlinear optimization problems. AD

Discrete Mathematics, T(14: 1). *Discrete Mathematics: An Introduction to Concepts, Methods, and Applications.* Jerrold W. Grossman. Macmillan, 1990, xxvi + 911 pp, $36.75. [ISBN: 0-02-348331-8] Treats elementary logic, set theory, functions and relations, algorithms, induction and recursion, combinatorics. More thorough treatment of graph theory than books at a comparable level. AD

Number Theory, P. *Diophantine Equations and Inequalities in Algebraic Number Fields.* Wang Yuan. Springer-Verlag, 1991, xvi + 168 pp, $69. [ISBN: 0-387-52019-8] A concise up-to-date survey of the field; topics include the circle method, exponential sums, Weyl's sums, singular series and integrals, Waring's problem, and inequalities for forms. SG

Group Theory, S(18), P. *Topics in Representation Theory.* Ed: A.A. Kirillov. Adv. in Soviet Math., V. 2. AMS, 1991, xi + 247 pp, $83. [ISBN: 0-8218-4101-7] An out-growth of the continuing seminar on Representation Theory at Moscow University: seven major papers on various topics by G.I. Olshankii, Y.A. Neretin, R.S. Ismagilov, A.I. Molev, O.D. Ovsienko, V. Yu. Ovsienko, and D.V. Yur'ev. Introduction; references. JS

Group Theory, P. *Arboreal Group Theory.* Ed: Roger C. Alperin. Math. Sci. Res. Instit. Public., V. 19. Springer-Verlag, 1991, x + 368 pp, $45. [ISBN: 0-387-97518-7] Proceedings of a conference held at MSRI in 1988. Focuses on groups acting on trees and applications to hyperbolic geometry. AD

Algebra, S(18), P. *Quadratic and Hermitian Forms over Rings.* Max-Albert Knus. Grund. der Math. Wissenschaften, B. 294. Springer-Verlag, 1991, xi + 524 pp, $98. [ISBN: 0-387-52117-8] Utilizing recent advances in techniques, particularly algebraic topology and K-theory, this book "exposes the theory of quadratic and Hermitian forms in a very general setting." Topics include Azumaya algebras, Clifford algebras, stability and cancellation, polynomial rings, and Witt groups. Extensive bibliography, index. JS

Algebra, T(15-17: 1, 2), L. *Abstract Algebra with a Concrete Introduction.* John A. Beachy, William D. Blair. Prentice Hall, 1990, xxiv + 325 pp. [ISBN: 0-13-004425-3] To ease transition into abstraction, the text begins with leisurely chapters on number theory and function (including permutations) before defining groups in Chapter 3. Sophistication level rises with chapters on fields, group structure theory, and Galois theory. Unusual effort is made to help the student understand the nature of algebra and the importance and technique of proof. Exercises (no solutions), bibliography, index. JS

Calculus, S, C. *GyroGraphics.* (Version 3.0). Nancy Anne Johnson. CipherSystems (171 Willow Drive, Stillwater, OK 74075; 405-377-4432), 1990, $75. An effective graphing tool for IBM-PC compatible computers that makes it easy to create two- and three-dimensional graphs in a number of forms (rotated, polar, cylindrical, spherical, parametric, etc.); to rotate them stepwise or continuously; to zoom and unzoom; and to examine vector fields as well as tangent and acceleration vectors in easily sequenced "layers." Good selection of examples in an opening menu; straightforward options to input user-defined

functions. Requires graphics board; performance improves significantly with greater machine power. LAS

Calculus, T(13: 1, 2). *Single-Variable Calculus With Discrete Mathematics.* John Feroe, Charles Steinhorn. Harcourt Brace Jovanovich, 1991, xii + 1021 pp, $38 net. [ISBN: 0-15-581062-6] An evolutionary rather than revolutionary calculus reform textbook. Ideas from discrete mathematics are presented near the beginning and used throughout the text (e.g., limits are introduced via sequences). The last four chapters treat topics that explore relationships between discrete and continuous mathematics: generating functions, probability, and difference and differential equations. AO

Real Analysis, T(15-16), S, L. *Standard and Nonstandard Analysis: Fundamental Theory, Techniques and Applications.* R.F. Hoskins. Math. & Its Applic. Ellis Horwood (US Distr: Prentice Hall), 1990, 269 pp, $65.95. [ISBN: 0-13-840844-0] The first book to introduce the basic ideas of nonstandard analysis along-side standard topics in analysis. Thus, the book presents a standard introduction to analysis (number systems, convergence, functions, continuity, differentiability, integration), but at the same time introduces nonstandard concepts at appropriate junctures. Little more than a working knowledge of elementary calculus is required, and the text makes no direct use of any explicit form of the "transfer principle," which is the key to the application of nonstandard analysis in its full sense. The book could be used as a standard beginning analysis textbook, with the use of infinitesimals and other nonstandard ideas taken at face value simply as an aid to comprehension. LCL

Complex Analysis, T(17: 3), L. *Complex Variables for Scientists and Engineers, Second Edition.* John D. Paliouras, Douglas S. Meadows. Macmillan, 1990, xiii + 586 pp. [ISBN: 0-02-390561-1] Straightforward, minimal introduction to basics of complex analysis followed by applications to boundary value problems, aerodynamics, and Schwarz-Christoffel integral. Maximum modulus theorem and Liouville's theorem covered in final chapter on theoretical results. Good selection of problems of varying difficulty. Solid, but rather dry presentation. BH

Differential Equations, P. *Boundary Value Problems for Operator Differential Equations.* V.I. Gorbachuk, M.L. Gorbachuk. Math. & Its Applic., V. 48. Kluwer Academic, 1991, xi + 347 pp, $149. [ISBN: 0-7923-0381-4] A study of equations of the form $y''(t) + A(t)y(t) = 0$ $(t\epsilon[a,b])$, where $A(t)$ is a semi-bounded, self-adjoint operator on a separable Hilbert space. Contains recent development of spectral analysis of the Sturm-Liouville operator differential equation (with an infinite-dimensional space and unbounded potential $A(t)$). MLR

Differential Equations, P. *Mathematical and Numerical Aspects of Wave Propagation Phenomena.* Eds: Gary Cohen, Laurence Halpern, Patrick Joly. SIAM, 1991, xiv + 794 pp, $68 (P). [ISBN: 0-89871-276-9] Proceedings of the first international conference on mathematical and numerical aspects of wave propagation phenomena held in Strasbourg, France, April 1991. MLR

Differential Equations, T(14-15: 1-3), L. *Differential Equations, Theory and Applications.* Ray Redheffer, Dan Port. Jones & Bartlett, 1991, xiv + 722 pp, $48.75. [ISBN: 0-86720-200-9] A straightforward, well-written text with a variety of interesting applications. Topics range from the usual elementary differential equations to the more advanced Sturm-Liouville theory and nonlinear problems. MLR

Numerical Analysis, P. *The Numerical Solution of Two-Point Boundary Problems in Ordinary Differential Equations.* L. Fox. Dover, 1990, xi + 371 pp, $9.95 (P). [ISBN: 0-486-66495-3] Reprint of the 1957 edition first published by Clarendon Press. Detailed coverage of numerical methods for boundary value and eigenvalue problems in ordinary differential equations of order up to four, including appropriate treatment of different boundary conditions. Includes also a chapter on the solution of linear systems corresponding to above methods, and a discussion concerning the difference between the methods for the initial value problems and the boundary value problems. LS

Numerical Analysis, T(13: 3). *Numerical Mathematics.* Günther Hämmerlin, Karl-Heinz Hoffmann. Transl: Larry Schumaker. Undergrad. Texts in Math. Springer-Verlag, 1991, xi + 422 pp, $39.95 (P). [ISBN: 0-387-97494-6] Covers the usual classical topics of numerical analysis, but also includes recent developments such as up-to-date treatments of splines, multidimensional analysis, linear optimization methods, and Monte-Carlo methods. Emphasizes numerical insights as well as mathematical vigor. LS

Operator Theory, T(18), P. *Operator Theory in Function Spaces.* Kehe Zhu. Pure & Appl. Math., V. 139. Marcel Dekker, 1990, ix + 258 pp. [ISBN: 0-8247-8411-1] Toeplitz, Hankel, and composition operators on Bergman and Hardy spaces of the unit disk. Main emphasis is on boundedness, compactness, and membership in the Schatten classes. Minimal prerequisites. MLR

Functional Analysis, P. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems.* Ioana Cioranescu. Math. & Its Applic., V. 62. Kluwer Academic, 1990, xiv + 260 pp, $99. [ISBN: 0-7923-0910-3] A duality-map approach to nonlinear operators related to nonlinear functional and evolution equations. Develops the necessary convex analysis and Banach-space geometry. MLR

Analysis, P. *Variations on a Theme by Kepler.* Victor Guillemin, Shlomo Sternberg. Colloq. Public., V. 42. AMS, 1990, vii + 88 pp, $43. [ISBN: 0-8218-1042-1] The authors "touch on the mysterious role that groups, especially Lie groups, play in revealing the laws of nature." Kepler didn't know it, but his second law (equal areas in equal times) is due to the orthogonal group $O(3)$, while the first and third

stem from $O(4)$. For that matter, "even the symplectic groups in eight- and twelve-dimensional space get involved." BC

Analysis, T(18), S, P, L. *Analysis IV: Linear and Boundary Integral Equations.* Eds: V.G. Maz'ya, S.M. Nikol'skiĭ. Encyclop. of Math. Sci., V. 27. Springer-Verlag, 1991, 233 pp, $59. [ISBN: 0-387-51997-1] First part contains survey of current state of theory of linear integral equations, presenting necessary abstract operator theory before discussing, in detail, Fredholm operators. Latter sections investigate one- and higher-dimensional singular operators. Second part concentrates on theoretical aspects of boundary integral equation methods, giving survey of classical theory of harmonic potentials. Includes applications to problems of mathematical physics. KS

Analysis, T(18), P. *Lecture Notes in Mathematics-1461: Abel Integral Equations.* Rudolf Gorenflo, Sergio Vessella. Springer-Verlag, 1991, vii + 215 pp, $25 (P). [ISBN: 0-387-53668-X] Elementary theory and many applications are discussed in the first part. Harder topics, such as the behavior of Abel integral operators in function spaces, are also presented. MLR

Differential Geometry, S(18), P. *Lie Groups I-II.* Mikio Ise, Masaru Takeuchi. Transl. of Math. Mono., V. 85. AMS, 1991, iv + 259 pp, $89. [ISBN: 0-8218-4544-6] Translations of lecture notes. *Part I,* originally done by Ise, is a fast-paced introduction to Lie groups and algebras, concluding with classification and structure theory. *Part II,* by Takeuchi, discusses Riemannian and Hermitian symmetric spaces, leading to the Cartan classification theory. References, problems for *Part II.* JS

Differential Geometry, P. *Symplectic Geometry, Groupoids, and Integrable Systems.* Eds: P. Dazord, A. Weinstein. Math. Sci. Res. Instit. Public., V. 20. Springer-Verlag, 1991, xi + 311 pp, $48. [ISBN: 0-387-97526-8] Contains twenty papers from the Séminaire Sud Rhodanien de Géométrie organized at MSRI in 1989. Eight papers are in French. OJ

Differential Geometry, P. *Theory of Singularities and its Applications.* Ed: V.I. Arnold. Adv. in Soviet Math., V. 1. AMS, 1990, ix + 333 pp, $125. [ISBN: 0-8218-4100-9] Contains seventeen papers by participants of the singularities seminar at Moscow University. Translated from Russian. OJ

Geometry, P. *Geometry's Future.* Ed: Joseph Malkevitch. COMAP, 1991, ix + 105 pp, (P). Papers and recommendations from a small workshop convened by COMAP to explore new approaches to geometry: visualization, robotics, convexity, "geometry and the imagination," volumes, high school teachers. Authors T. Banchoff, D. Crowe, V. Klee, J. Malkevitch, W. Meyer, M. Senechal, and W. Thurston plan an innovative text as a sequel. LAS

Geometry, S, C. *Poincare: An Excursion into Hyperbolic Geometry, User's Manual.* (Version 3.01). George D. Parker. 1990, 18 pp, $50 site license; $20 individual. A menu-driven program for IBM-PC compatible computers that enable students to explore hyperbolic geometry. Primary data (points, lines, and constants) can be entered either directly or indirectly (via computation or by graphing); menu options use these data to graph and calculate relations involving lines, triangles, angles, perpendiculars, polygons, pencils, cycles, congruences, and more. Easy to use; includes documentation file. Requires graphics board; operates in either color or monochrome; alternate version (on source disk) for machines with math coprocessor. LAS

Algebraic Topology, T(17: 2), P. *A Basic Course in Algebraic Topology.* William S. Massey. Grad. Texts in Math, V. 127. Springer-Verlag, 1991, xvi + 428 pp, $49.95. [ISBN: 0-387-97430-X] Combines most of two previous books into one volume. The first five chapters—from *Algebraic Topology: An Introduction*—cover the classification of surfaces, the fundamental group, and covering spaces (*First Edition,* TR, November 1967; *Second Edition,* TR, April 1978). The next ten chapters—from *Singular Homology Theory*—cover the standard topics in homology and cohomology using cubes rather than simplices to define singular homology (TR, June-July 1981). Geometric motivation of ideas, historical comments, applications to geometric/topological problems, and engaging exposition make this an attractive text. Exercises. OJ

Topology, S(17). *General Topology I: Basic Concepts and Constructions, Dimension Theory.* Eds: A.V. Arkhangelskiĭ, L.S. Pontryagin. Encyclop. of Math. Sci., V. 17. Springer-Verlag, 1990, 202 pp, $59. [ISBN: 0-387-18178-4] Translated from the Russian. First half gives a densely written, rather dry introduction to the basics of topology which assumes a rather high level of mathematical sophistication on the part of the reader. Second half presents some of the more important concepts in dimension theory, and includes a nice historical introduction. No exercises. BH

Dynamical Systems, S, P, L. *Fractals: Non-Integral Dimensions and Applications.* Ed: G. Cherbit. Transl: F. Jellett. Wiley, 1991, xv + 249 pp, $69.95. [ISBN: 0-471-92798-8] Translation of the 1987 Masson (Paris) *Fractals: Dimensions nonentières et applications* containing papers from the 1985-86 Hausdorff Seminar sponsored by the biophysics research group of the University of Paris. Papers are generally expository, covering a wide range of related topics (stochastic processes, dimension, biometry, entropy, etc.). Includes a reflective personal commentary by Mandelbrot, and a brief preface by J.-P. Kahane. LAS

Control Theory, P. *Lecture Notes in Control and Information Sciences-150: Robust Stabilization in the Gap-topology.* L.C.G.J.M. Habets. Springer-Verlag, 1991, vii + 126 pp, $25 (P). [ISBN: 0-387-53466-0] Theoretical treatment of stabilization, with algorithms that make it practical. BC

Stochastic Processes, P. *Excessive Measures.*

R.K. Getoor. Prob. & Its Applic. Birkhauser Boston, 1990, ix + 189 pp, $39. [ISBN: 0-8176-3492-4] The author presents a unified treatment of recent results on the potential theory of excessive measures. Highly technical, but timely. TAV

Statistics, S(17-18), L. *The Engineering Statistician's Guide to Continuous Bivariate Distributions.* T.P. Hutchinson, C.D. Lai. Rumsby Scientific, 1991, xxii + 346 pp, $44 (P). [ISBN: 0-646-02413-2] This reference guide lists bivariate distributions, their properties, and how they have been used as models in the statistical and engineering literature. Marginals are generally transformed into some convenient standard form such as the uniform, normal, or exponential before specifically bivariate aspects are considered. Nearly fifty pages of references, most of these to literature published in the last few years. RWJ

Statistics, L, P*. *Spatial Statistics and Digital Image Analysis.* National Research Council. National Academy Pr, 1991, ix + 234 pp, (P). [ISBN: 0-309-04376-X] Ten expository essays introducing the growing field of spatial statistics—the study of spatially referenced data. Primary applications are to image analysis (e.g., in satellite reconnaissance of missle silos or wheat crops, in tomography, in atmospheric and oceanographic studies), with spill-over to ecology, stereology, speech recognition, and other types of pattern analysis. Many applications are driven by computation (to find patterns in data) and applied to environmental issues. A timely invitation to a field of rapidly increasing importance. LAS

Statistics, P. *Sample Size Methodology.* M.M. Desu, D. Raghavarao. Stat. Modeling & Decision Sci. Academic Pr, 1990, xii + 135 pp, $32.50. [ISBN: 0-12-212165-1] Summarizes sample sizes needed to meet error probability, power, and variance constraints for various estimation and hypothesis testing problems. Includes ranking and selection problems, design of clinical trials, and sampling from finite populations. Relevant tables referenced in Appendix. Contains some problems. RWJ

Statistics, S, C**, P*.** *MacSpin: 3-Dimensional Data Analysis Software.* Abacus Concepts (1984 Bonita Ave., Berkeley, CA 94704; 415-540-1949), 1989, 207 pp, (P). *Version 3.0 Release Notes,* 1991, ii + 52 pp, (P); $295; site licenses available. A powerful Macintosh software package for exploring multidimensional data sets by displaying rotating "clouds" of three-dimensional data, with a fourth dimension represented by slices controlled by a slider that can be moved continuously for animation. Standard Macintosh interface tools provide options for entering data (including cut-and-paste from spreadsheets and other sources), for transforming variables, and for printing various versions of labelled graphs. Pop-up windows give statistical summaries of variables. Version 3.0 also includes standard tools for two-dimensional charts interfaced to the MacSpin data sets; "overlays" to crop, move, resize, and add captions to graphs produced by MacSpin; variable scaling; and a MacPaint-style graphic tool palette. LAS

Computer Literacy, S*, P*, L*. *Writing Space: The Computer, Hypertext, and the History of Writing.* Jay David Bolter. Lawrence Erlbaum Assoc, 1991, xiii + 258 pp, $22.50 (P); $45. [ISBN: 0-8058-0428-5; 0-8058-0427-7] A compelling, scholarly *apologia* for the new paradigm of electronic (hypertext) literacy where writing is "open to performance by the reader." "Writing itself is not merely influenced by technology, but rather *is* technology." Classicist Bolter analyzes today's revolution in text in terms of the whole history of writing, looking at visual writing space, conceptual writing space, and the mind as a writing space. Touches on a wide range of users from deconstructionism to cultural literacy, from Homeric oral traditions to artificial intelligence. A superb reading for introductory courses in computer literacy. LAS

Elementary Computer Science, P, L. *The Age of Intelligent Machines.* Raymond Kurzweil. MIT Pr, 1990, xiii + 565 pp, $39.95. [ISBN: 0-262-11121-7] A grand tour of the past, present, and future of computers and computer science. Filled with photographs, graphs, and illustrations. Accessible to non-specialists, but interesting to specialists. Includes articles by Minsky, Dennett, Turkle, Hofstadter, Papert, and more. Topics range from philosophical roots to mechanical roots, from artificial intelligence to the science of art. MK

Languages, S, P, L*. *Commentary on Standard ML.* Robin Milner, Mads Tofte. MIT Pr, 1991, 160 pp, $19.95 (P). [ISBN: 0-262-63137-7] A companion volume to *Definition of Standard ML* by Milner, Tofte, and Harper. Gives an in-depth explanation of the semantic theory of *ML*. A notable feature is the inclusion of information about *why* certain design decisions were made. AO

Languages, T(16-17: 1), L*. *Introduction to the Theory of Programming Languages.* Bertrand Meyer. Intern. Ser. in Comput. Sci. Prentice Hall, 1990, xvi + 447 pp. [ISBN: 0-13-498510-9] A modern, formal treatment of mathematical models of programming languages. The topics covered include syntax, especially abstract syntax, denotational and axiomatic semantics, and the complementarity of these two approaches to programming language semantics. AO

Algorithms, T(16-17: 1, 2), L*. *Algorithms from P to NP, Volume 1: Design & Efficiency.* B.M.E. Moret, H.D. Shapiro. Benjamin/Cummings, 1991, xv + 576 pp, $38.75. [ISBN: 0-8053-8008-6] Discusses algorithm design techniques and illustrates these techniques by applying them to the design of efficient algorithms for some well-known problems in combinatorial optimization. Algorithms are given with proofs of correctness, an analysis of their complexity, and a discussion of implementation details (often including an empirical study of competing designs). AO

Computer Systems, T(13). *Programming and Designing with the 68000 Family.* Tibet Mimar. Prentice Hall, 1991, xiii + 641 pp. [ISBN: 0-13-

731498-1] Intended to provide a thorough introduction to the hardware and software architecture of the Motorola 68000 family of processors, including the 68010, 68020, and 68030. It covers the 68000 series architecture, instruction set, interrupt handling, memory system, and coprocessor design. GMS

Computer Systems, S*, C, P, L*. *Exploring Mathematics with Mathematica: Dialogs Concerning Computers and Mathematics.* Theodore W. Gray, Jerry Glynn. Addison-Wesley, 1991, 535 pp. [ISBN: 0-201-52809-6] An eclectic, chatty tour of the computer algebra system *Mathematica* showing off its power with applications both traditional (e.g., iteration, fractals, graphics, Fourier transforms) and bizarre (mathematical transformations of Beethoven's Ninth Symphony). Text is written as a glib rhetorical dialogue; topics illustrate music, graphics, and school mathematics. The entire text was prepared as a *Mathematica* Notebook—text, sound, graphics—and is provided electronically on a CD-Rom inside the back cover, thus enabling *Mathematica* users to interact directly with the examples by running and modifying them. LAS

Computer Systems, T(15-16), P, L. *An Introduction to Text Processing.* Peter D. Smith. MIT Pr, 1990, xii + 300 pp, $32.50. [ISBN: 0-262-19299-3] A survey of past and present practice in computer approaches to handling natural language text: storage, editors, compression, encryption, formatting, concordances, hyphenation, spelling checkers, syntax analysis, authorship studies, translation. Major examples come from UNIX tools; many references to other packages. Exercises and projects make the book suitable as a text. LAS

Computer Systems, P. *Application Environment Specification (AES) User Environment Volume, Revision B.* Open Software Foundation. Prentice Hall, 1991, xii + 1131 pp. [ISBN: 0-13-043530-9] A detailed specification of the OSF's Application Environment Specification (AES). The bulk of this document is made up of the manual pages for all of the AES User Environment interfaces. AO

Computer Systems, S*, P. *UNIX System V.4 Command Summary.* SSC, 1990, 80 pp, $8 (P). [ISBN: 0-916151-44-1] A concise envelope-size pamphlet with summaries of syntax and flags for most UNIX commands that are accessible to ordinary users. Includes expanded sections on use of tools such as `ftp`, `sdb`, `sed`, and `telnet`. LAS

Computer Graphics, S*, P, L.** *The Second Computer Revolution: Visualization.* Richard Mark Friedhoff, William Benzon. WH Freeman, 1989, 215 pp, (P). [ISBN: 0-7167-2231-3] An extraordinary exposition of vision and images, profusely illustrated in full color, beginning with Edwin Land's revolutionary "retinex" theory of color vision and proceedings to the most recent computer-created artificial realities. Covers image enhancement, object-based images, procedural (including fractal) methods, animation, CAD systems, and much more. Each topic is explained clearly in simple concise terms, and illustrated with apt visual images. A remarkably interdisciplinary endeavor, both well-written and vividly illustrated. LAS

Computer Science, T(16-17: 1), S, P, L. *Genetic Algorithms in Search, Optimization, and Machine Learning.* David E. Goldberg. Addison-Wesley, 1989, xiii + 412 pp. [ISBN: 0-201-15767-5] A user-friendly introduction to genetic algorithms: computer search methods based on principles of natural selection and genetics. Includes problems and computer assignments. Suitable as a text or for self-study. BC

Computer Science, T(16-18: 1), P, L.** *Mathematics for the Analysis of Algorithms, Third Edition.* Daniel H. Greene, Donald E. Knuth. Progress in Comput. Sci. & Appl. Logic, V. 1. Birkhauser Boston, 1990, viii + 132 pp, $24.95. [ISBN: 0-8176-3515-7] This edition includes some improvements in the exposition, minor corrections, and new appendices including examination problems (and solutions) from 1982 and 1988 (*First Edition*, TR, August-September 1982). Based on materials developed for an advanced computer science course at Stanford University. Presumes that the reader is familiar with the fundamentals of complex variable theory and combinatorial analysis. AO

Computer Science, S(15-16), P*, L*. *ACM Turing Award Lectures: The First Twenty Years, 1966-1985.* ACM. Anthology Ser. Addison-Wesley, 1987, xviii + 483 pp, (P). [ISBN: 0-201-54885-2] These lectures by the first 23 winners of the Turing Award give a wide-ranging perspective on the past, present, and possible future of computer science. A valuable set of essays. BC

Computer Science, S, P, L*. *Technology 2001: The Future of Computing and Communications.* Ed: Derek Leebaert. MIT Pr, 1991, xvi + 392 pp, $29.95. [ISBN: 0-262-12150-6] Fourteen essays by leaders and visionaries in computer R&D on the next ten years of computing: microprocessors, supercomputing, networks, electronic documents, telecommunications, imaging, public policy—all replete with implications for education. A thought-provoking glimpse into the world today's students will inherit. LAS

Applications, T(15: 2), P, L. *Applied Analysis by the Hilbert Space Method: An Introduction with Applications to the Wave, Heat, and Schrödinger Equations.* Samuel S. Holland, Jr. Pure & Appl. Math., V. 137. Marcel Dekker, 1990, xii + 560 pp, $55. [ISBN: 0-8247-8259-3] Well-written text for a one-year undergraduate applied mathematics course. Prerequisites are calculus (three semesters), linear algebra, and basic physics. Includes necessary sections on differential equations. Gets to the applications quickly. MLR

Applications, P*. *Walsh Series and Transforms: Theory and Applications.* B. Golubov, A. Efimov, V. Skvortsov. Math. & Its Applic., V. 64. Kluwer Academic, 1991, xiii + 368 pp, $169. [ISBN: 0-7923-1100-0] Translated from a 1987 Russian work. An account of the theoretical foundations of Walsh series

and transforms. For engineers, technical specialists, and applied mathematicians. Applications to digital information processing, digital filtering, and digital holograms. Appendices with background information on group theory, measure theory, the Lebesgue integral, and functional analysis. Brief commentaries and historical references for each chapter. Extensive list of references. JDEK

Applications (Behavioral Science), S, L*. *Negotiation Games: Applying Game Theory to Bargaining and Arbitration.* Steven J. Brams. Routledge Chapman and Hall, 1990, xviii + 297 pp, $15.95 (P); $49.50. [ISBN: 0-415-90338-6; 0-415-90337-8] A model-based (not a mathematics-based) survey of negotiations (e.g., Cain vs. God, Sadat vs. Begin, Kennedy vs. Khruschev), rooted in game theory, but extended to common systems of bargaining and arbitration. Chief message: the impediment to resolving conflicts is often rationality, not irrationality. LAS

Applications (Biological Science), T(15-17), S, P, L*. *The Geometry of Biological Time.* Arthur T. Winfree. Study Edition. Springer-Verlag, 1990, xiv + 530 pp, $32 (P). [ISBN: 0-387-52528-9] Corrected paperback reprint of the 1980 original (hardcover) edition (TR, December 1980). Diverse biological applications of periodic dynamics: circadian cycles, energy metabolism, Mitotic cycles, waves, growth, and much more in 23 chapters that employ a wide range of methods from undergraduate mathematics. Massive bibliography (only through 1978) documents the roots of this very rapidly growing field. LAS

Applications (Physical Science), S(17-18), P. *Flux Coordinates and Magnetic Field Structure: A Guide to a Fundamental Tool of Plasma Theory.* W.D. D'haeseleer, *et al.* Ser. in Computat. Physics. Springer-Verlag, 1991, xii + 241 pp, $98. [ISBN: 0-387-52419-3] For non-specialists as well as established researchers in the fields of plasmas and controlled fusion. Systematic and rigorous treatment of the mathematical framework of magnetic-flux-coordinate systems (in which magnetic field lines appear as straight lines). Focus is on the geometrical properties of vectors and tensors in curvilinear coordinates. JDEK

Applications (Physical Science), P. *Spectral Theory of Families of Self-Adjoint Operators.* Y.S. Samoilenko. Math. & Its Applic., V. 57. Kluwer Academic, 1991, xvi + 293 pp, $129. [ISBN: 0-7923-0703-8] Studies the mathematical models of physical systems; presents the connected theory of countable families of self-adjoint operators on a Hilbert space and their spectral properties; deals with countable collections of self-adjoint operators and their joint spectral properties; illustrates connection with Lie groups. LS

Applications (Physical Science), P. *Continuum Models and Discrete Systems.* Ed: G.A. Maugin. Interaction of Mech. & Math. Ser. Longman Scientific & Tech (US Distr: Wiley), 1990, $150 each, [ISBN: 0-470-21592-5]. *Volume 1,* 324 pp; *Volume 2,* 324 pp. Texts from most of the short papers presented at an international symposium (CMDS6) held in June 1989 at the University of Bourgogne in Dijon, France. Papers in each volume are grouped in four parts: complex fluids; deformable solids; microstructure, thermodynamics, and geometry; and nonlinear excitations. LAS

Applications (Physical Science), P. *Dynamics of Polyatomic Van der Waals Complexes.* Eds: Nadine Halberstadt, Kenneth C. Janda. NATO ASI Ser. B., V. 227. Plenum Pr, 1990, ix + 542 pp, $120. [ISBN: 0-306-43612-4] Thirty-nine papers from a NATO workshop on theoretical and experimental studies of weakly bound molecular complexes. Focuses on extending current knowledge of triatomic systems to polyatomic molecules and clusters. BC

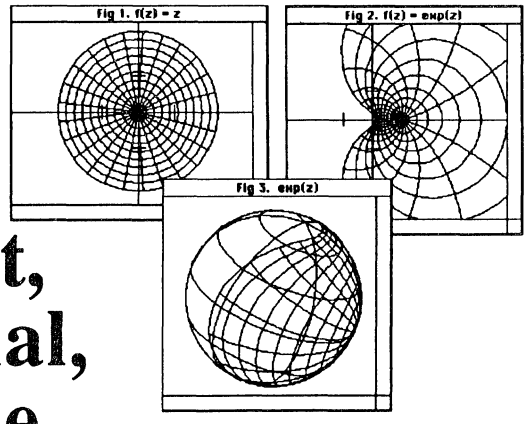
Applications (Physics), S*(15-16). *Problems in Theoretical Physics, Volumes I-II.* Willi-Hans Steeb. Bibliographisches Institut, 1990. *Volume I: Introductory Problems,* 197 pp, (P), [ISBN: 3-411-14451-3]; *Volume II: Advanced Problems,* 290 pp, (P). [ISBN: 3-411-14611-7] *Volume I* contains 120 problems on topics which should be familiar to undergraduates having completed courses in linear algebra and complex analysis. Whenever appropriate, connections to physics are cited. *Volume II* contains 136 advanced problems in ordinary and partial differential equations, vector and tensor analysis, differential forms, generalized functions, quantum theory and gauge theory. Un-wordy and clear solutions are given for almost every problem. Both volumes are lightly sprinkled with helpful asides. Unfilled pages (created by insistence on beginning each problem at the top of a page) cry for more. JDEK

Applications (Physics), S. *Algebraic and Spectral Methods for Nonlinear Wave Equations.* N. Asano, Y. Kato. Pitman Mono. & Surveys in Pure & Appl. Math., V. 49. Longman Scientific & Tech (US Distr: Wiley), 1990, xvii + 418 pp, $158. [ISBN: 0-470-21417-1] Presents algebraic and spectral methods for nonlinear wave equations or nonlinear partial differential equations; introduces the method of continuous groups, the method of inverse scattering formulation, and fundamental materials of the Lie's transformation method; covers the development and new results in the last twenty years in these and related fields. LS

Reviewers

RJA: Richard J. Allen, St. Olaf; RB: Richard Brown, St. Olaf; BC: Barry Cipra, St. Olaf; AD: Amy Davidow, Macalester; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; BL: Brian Loe, Carleton; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LS: Li Sheng, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf.

Software and video tapes for the student, professional, and anyone who loves mathematics.



Ask for our latest catalog.

Lascaux Graphics 3220 Steuben Ave.
Bronx, NY 10467 USA (800) 338-0993

Theory for Practice

Introduction to Wavelets
Charles K. Chui
August 1991, c. 250 pp.
$49.95 (tentative)
ISBN: 0-12-174584-8

Elements of Scientific Computing and Differential Equations
An Introduction to Numerical Methods
SECOND EDITION
Gene Golub and James M. Ortega
September 1991, c. 400 pp.
$49.95 (tentative)
ISBN: 0-12-289255-0

Curves and Surfaces
edited by
Pierre-Jean Laurent, Alain Le Mehaute, and Larry L. Schumaker
July 1991, c. 536 pp., $49.95
ISBN: 0-12-438660-1

Differential Equations and Mathematical Physics
Christer Bennewitz
July 1991, 400 pp., $49.95
ISBN: 0-12-089040-2

Nonlinear Equations in the Applied Sciences
W.F. Ames and Colin Rogers
July 1991, 512 pp., $49.50
ISBN: 0-12-056752-0

Order from your local bookseller or directly from

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Department #01091
1250 Sixth Avenue, San Diego, CA 92101
CALL TOLL FREE
1-800-321-5068
FAX: **1-800-235-0256**
Quote this reference number for *free postage and handling* on your prepaid order ➥ **01091**

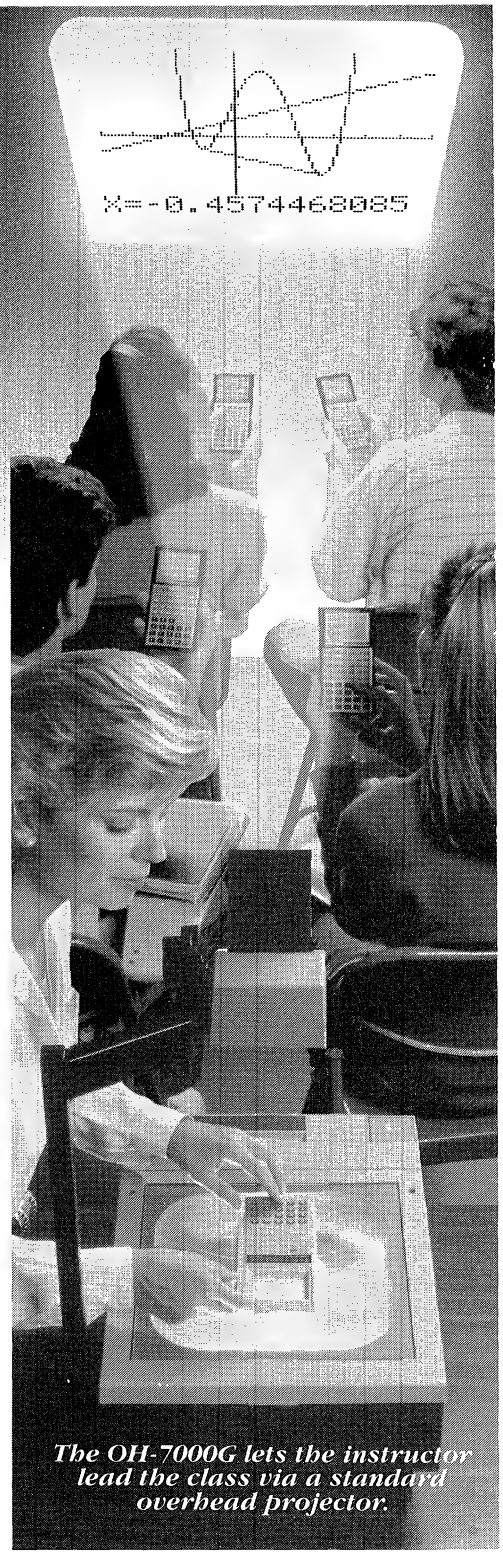
Prices subject to change without notice
©1991 by Academic Press, Inc.
All Rights Reserved LH/SS—01091

The Numerical Method of Lines
Integration of Partial Differential Equations
W.E. Schiesser
July 1991, 326 pp., $69.95
ISBN: 0-12-624130-9

Markov Processes
An Introduction for Physical Scientists
Daniel T. Gillespie
September 1991, c. 550 pp.
$49.95 (tentative)
ISBN: 0-12-283955-2

Misteaks and How to Find Them Before the Teacher Does
A Calculus Supplement
SECOND EDITION
Barry Cipra
1989, 65 pp., $7.95
ISBN: 0-12-174695-X

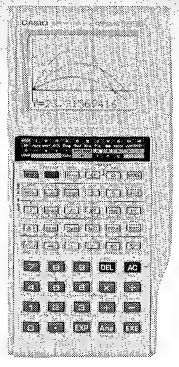
The Power to Project Your Ideas



The OH-7000G lets the instructor lead the class via a standard overhead projector.

Introducing the OH-7000G Projectable Graphic Calculator.

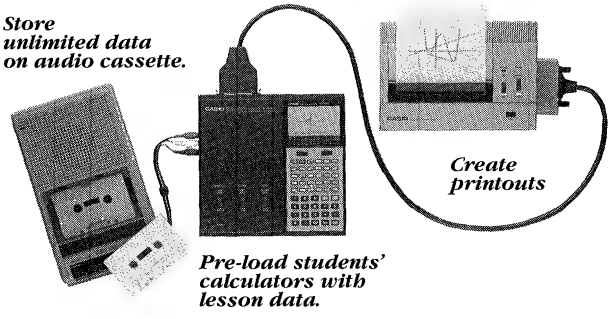
Casio has taken one of its most powerful graphic calculators and turned it into the most powerful **teaching tool** of its kind. The OH-7000G features a transparent, projectable screen so that teachers can lead their class just by using a standard overhead projector.



OH-7000G

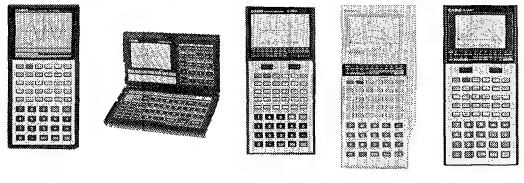
Since inventing the graphic calculator in 1985, Casio has created an entire family of exciting classroom products. Five different graphic calculator models are available, each offering **hundreds of powerful, built-in functions**.

With our **printers**, you can create "hard copy" of graphs and calculations for photocopying onto test papers or transparencies. The **FA-80 Data Interface** lets you store unlimited data on audio cassettes.



The possibilities are endless. That's why the **NCTM curriculum** calls for the use of graphic calculators. Many **new textbooks** are being published especially for use with this technology.

So, take a bold, new step and enter Casio's "Electronic Classroom" – a place where the solution is always clear.



CASIO

Creators of the Graphic Calculator

570 Mt. Pleasant Avenue, Dover, New Jersey 078(

Casio Educational Product Distributors

ABC SCHOOL SUPPLY
800-669-4ABC
(IN GA 404-497-0001)

ALLIED NATIONAL
800-999-8099
(IN MI 313-543-1232)

ARROWHEAD BUSINESS MACHINES
800-234-3396
(IN TX 214-869-0721)
(IN MO 816-861-1113)
(IN CA 213-946-6680)

THE BACH COMPANY
800-248-2224
(IN CA 415-424-0800)

BARDEEN SCHOOL SUPPLY
800-831-0113
(IN NY 315-437-7566)

BECKLEY-CARDY COMPANY
800-446-1477

BHARDS PUBLISHING
800-473-7999
(IN IL 312-642-8657)

CALCULATORS INC.
800-533-9921
(IN MN 800-533-9921)

CAROLINA WHOLESALE
800-521-4600
(IN NC 704-598-8101)

COLBORN SCHOOL SUPPLY INC.
800-548-7031
(IN CO 303-778-1220)
(IN MT 406-245-3158)

COPCO
800-446-7021
(IN OH 800-589-3006)

CUISENAIRE
800-237-3142
(IN NY 914-235-0900)

DALE SEYMOUR PUBLICATIONS
800-872-1100
(IN CA 800-872-1100)

THE DOUGLAS STEWART CO.
800-279-2795
(IN WI 800-279-2795)

E.A.I.
800-272-0272
(IN NJ 201-891-9466)

EDUCATIONAL ELECTRONICS
800-526-9060
(IN MA 617-821-6458)

GLOBAL PRODUCTS
800-633-0633
(IN IL 708-397-4944)

GREYSTONE EDUCATIONAL MATERIALS
800-733-0671
(IN MN 612-430-9857)

KURTZ BROTHERS
800-252-3811
(IN PA 814-765-6561)

LONGINO DISTRIBUTORS
800-633-6224
(IN NC 704-873-3282)

NASCO
800-558-9595
(IN WI 414-563-2446)

NATIONAL AUDIO-VISUAL SUPPLY
800-222-0109
(IN NJ 800-222-0109)

NEMESIS DISTRIBUTING
800-940-7407
(IN FL 305-477-8822)

PENNS VALLEY PUBLISHING/ LEARNING SYSTEMS
800-422-4412
(IN PA 215-855-4948)

SERVCO PACIFIC
(IN HI 808-841-7566)

TAM'S
800-421-5188
(IN CA 800-244-5624)

TAYLOR ELECTRIC
800-558-6970
(IN WI 800-242-8940)

TECHLINE
800-777-3635
(IN VA 703-389-0857)

TROXELL COMMUNICATIONS INC.
800-528-7912
(IN AZ 800-352-7941)

VALLEY BUSINESS MACHINES
(IN UT 801-969-6303)

VISTATECH
800-847-9851
(IN NY 212-254-9851)
(IN CA 213-602-0277)

WHOLESALE ELECTRONIC SUPPLY
800-527-2156
(IN TX 800-441-0145)

CASIO®

Introducing E.Z. Math, E.Z. Algebra and E.Z. Arithmetic for the HP 48SX

E.Z. Math, E.Z. Algebra and E.Z. Arithmetic are programs for the Hewlett Packard 48SX calculator conceived, written and programmed by Raymond La Barbera and the E.Z. Software Company. Each program comes on a 128K plug-in ROM card accompanied by an easy-to-understand, well-written, detailed manual loaded with lots of specific examples and is designed for use by students, teachers. parents and business people. Each program features an easy-to-use, logically organized, user-friendly interface which enables those who consider themselves to be calculator and computer illiterates, as well as those who don't like to read manuals, to have full access to all program features quickly and easily. Since the HP 48SX is essentially an impressive, sharp-looking, 8 ounce pocket computer, students are easily motivated to take it along with them to study, practice, drill and master math in a study hall, on a train or bus, in a car, on line, on vacation, on a break—in short, for self-study at any time and in any place.

What Can Be Done With E.Z. Math

E.Z. Math effectively solves problems involving graphs, numbers, loans and savings. With E.Z. Math, anyone can:

- Master the entire high school and college graphing curriculum, from algebra to calculus, with 188 families of equations, inequalities, functions, and systems, all arranged in an easy-to-use, user-friendly system of menus to make graphic analysis a snap!
- Get extensive help with calculations involving fractions, whole numbers, complex numbers and number sequences.
- Easily do savings and loan calculations and generate complete amortization tables.
- Learn many basic concepts including those involving sets, variables, graphing, solving, numbers, loans and savings.

What Can Be Done With E.Z. Algebra

E.Z. Algebra is a comprehensive ninth grade high school basic algebra course as well as a high school and college remedial algebra course that builds a solid algebra foundation. With E.Z. Algebra, anyone can:

- Learn about sets, operations, variables, relations and other concepts essential to a real understanding of algebra.
- Understand the sets of natural numbers, whole numbers, integers, rational numbers and real numbers.
- Master the meaning and properties of the operations of addition, subtraction, multiplication, division, power and root.
- Do all kinds of problems involving algebraic expressions, numerical phrases, equations and inequalities.

What Can Be Done With E.Z. Arithmetic

E.Z. Arithmetic is a comprehensive elementary school basic arithmetic course as well as a high school and college remedial arithmetic course that makes solving most arithmetic problems a snap! With E.Z. Arithmetic, anyone can:

- Learn how to add, subtract, multiply, divide and order whole numbers, fractions, decimals, percents and integers.
- Master the meaning, terminology and conversion methods for whole numbers, fractions, decimals and percents.
- Drill and be graded on endlessly varied, randomly selected sets of problems involving whole numbers, fractions, decimals and integers, with the difficulty level, number of problems, operation and type of number user selectable.

How To Order Copies or Get Further Information

Each E.Z. Software program costs $130.00 ($125.00 retail, plus $5.00 shipping and handling). Take a 10% discount when ordering ten or more units. We accept payment by check, money order, COD, VISA, MC, AE and purchase order. If within 30 days you find that any E.Z. Software program fails to meet your expectations, we'll gladly take back your copy for a prompt, courteous refund. To order copies, either individually or bundled with HP 48SX calculators, please contact:

SMI Corporation, 250 West New Street, Dept MM2, Kingsport, Tennessee 37660
(800) 234-0123 or (615) 378-4821 or (615) 245-8982 (Fax).

NEW MEMBERS JOIN NOW NEW MEMBERS JOIN NOW

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036-1385

Name _____

Mailing Address _____

_____ Zip _____

Employer/School _____

Position (Rank) _____

Employer's City/State _____

Highest Earned Degree _____ Year Degree Earned _____

Institution Awarding Degree _____

Month/Year of Birth _____

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Find the column for your desired combination of MAA journals in the table below. All members receive FOCUS. **Select** the row appropriate for your initial membership period (1 year or 1.5 years) and your status (student or regular). **Circle** your dues in the table. **Write** the amount on the form below.

Journal codes used in the columns of the dues table are: THE AMERICAN MATHEMATICAL MONTHLY = **M**
MATHEMATICS MAGAZINE = **G** THE COLLEGE MATHEMATICS JOURNAL = **J**

Student Membership	**M**	**G**	**J**	**M + G**	**M + J**	**G + J**	**M + G + J**
1 year (Jan.–Dec. 1992)	$34.00	$28.00	$30.00	$ 40.00	$ 41.00	$ 35.00	$ 47.00
1.5 years (July 1991–Dec. 1992)	$50.00	$41.50	$44.50	$ 58.50	$ 60.50	$ 51.50	$ 69.00
Regular Membership							
1 year (Jan.–Dec. 1992)	$68.00	$55.00	$58.00	$ 82.00	$ 85.00	$ 72.00	$ 99.00
1.5 years (July 1991–Dec. 1992)	$99.50	$78.50	$83.50	$121.50	$126.50	$105.50	$148.50

These are **specially discounted rates for new members** provided to help the **MAA** reach a wider audience. They are not available to those who have been members for two years. **Student membership** is available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Student rates apply to unemployed persons who are seeking employment. **Annual dues include subscription prices as follows:** Regular Member $27 **M,** $17 **J,** $14 **G,** $6 **Focus.** Student member $11 **M,** $7 **J,** $5 **G,** $3 **Focus.** These rates are guaranteed for the indicated periods only.

Payment Enclosed $ _____ *U.S. Funds Only* (see below)

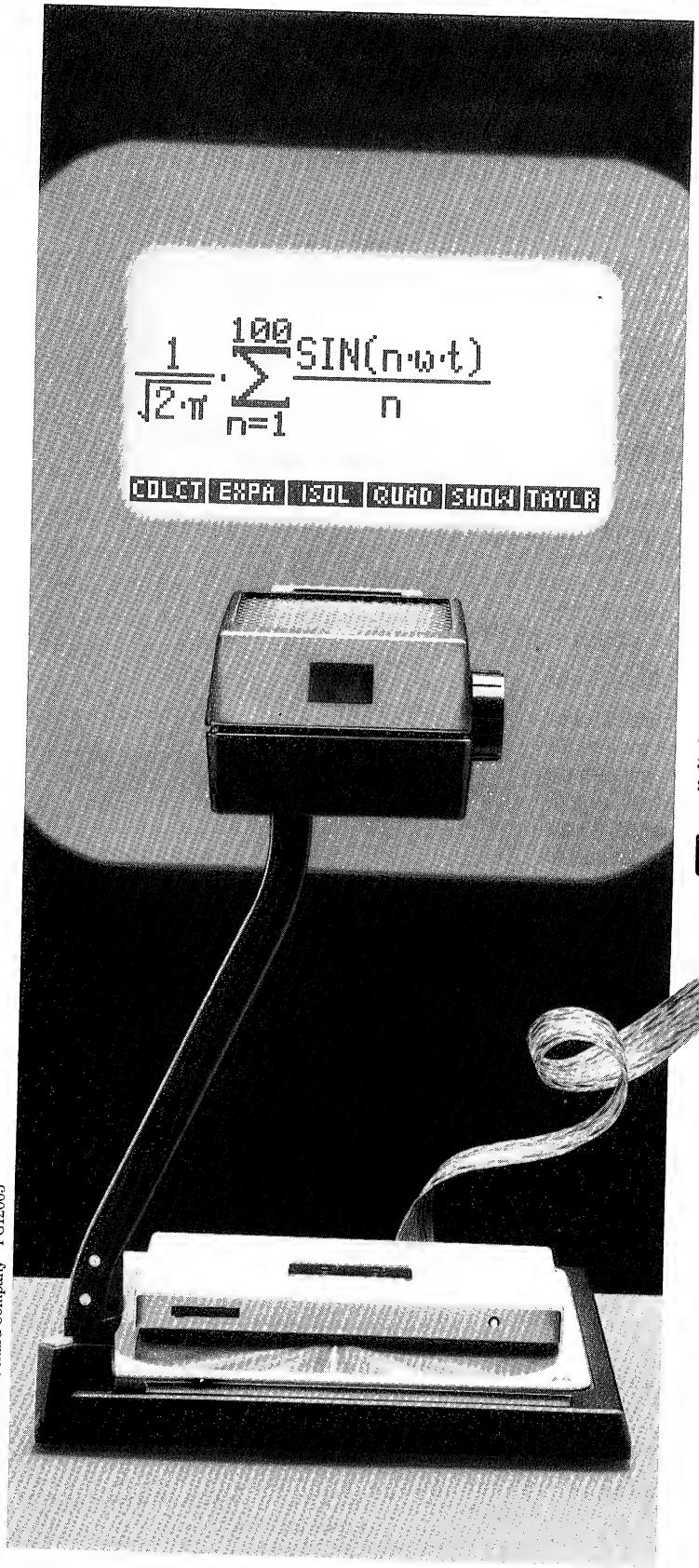
METHOD OF PAYMENT ☐ Check Enclosed ☐ VISA ☐ MasterCard

Card Number _____ Expiration Date |__|__|__|__|
(month/year)

Interbank Number |__|__|__|__| *(MasterCard only—located above name on card)*

SIGNATURE _____

Help your students discover more meaningful relationships.



Renewed for '91: a free classroom display device with purchase of 30 calculators.

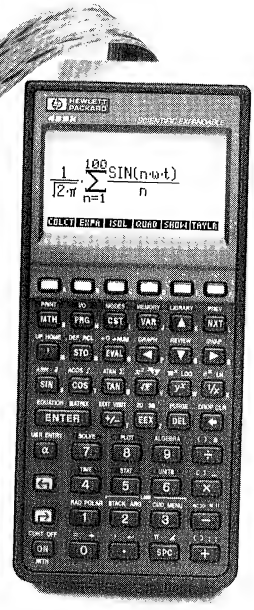
Showing is much more powerful than telling. So we've developed special classroom displays for our most advanced calculators.

The HP 48SX scientific expandable calculator, the cost-effective HP 28S and the new HP 48S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculators.

If your department or students purchase 30 HP 48SX, HP 28S or HP 48S calculators (or a mix of all three), we'll give you free an HP 48SX and plug-in classroom display (a $900 retail value). Or a classroom display version of the HP 28S (a $600 retail value). And free teaching materials.

Call **(503) 757-2004** from 8am to 3pm PDT for details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1991, and applies only to college and high school instructors.

HEWLETT PACKARD



© 1991 Hewlett-Packard Company PG12005

Winning Women into Mathematics

Patricia Clark Kenschaft, Editor

American media often ask why women "can't" do mathematics. Any answer is misleading. Better questions are needed, along with indications of how to find potential answers.

The Committee on the Participation of Women of the Mathematical Association of America was established in 1987 "to work for full involvement of women in MAA activities that will encourage women to pursue careers in the mathematical sciences." With this book, the Committee seeks to expand the number and effectiveness of those winning women into mathematics. WINNING WOMEN is written to inform, to empower, and to inspire.

The Committee identifies fifty-five cultural customs that discourage aspiring women mathematicians. They tell us how these customs can be changed and what can be done to recruit, retain, and acknowledge women in mathematics. A bibliography of over 100 sources on the issues of women's participation in mathematics is included, as well as descriptions of programs that have been successful in encouraging young women to study mathematics. The book is filled with interesting anecdotes, and contains over 50 photographs of prominent women in mathematics.

88 pp., 1991 , Paperbound
ISBN 0-88385-453-8

List: $11.00 MAA Member: $9.00

Catalog Number: WIW

CONTENTS

- A bibliography of over 100 sources on the issues of women's participation in mathematics
- Fifty-five cultural patterns causing American women to be underrepresented in mathematics
- What you personally can do
- Programs that succeed
- A history of women in mathematics-especially in the MAA
- A chronicle of the programs, articles, and suggestions of the Committee on the Participation of Women
- A minority woman's viewpoint
- An overview of the statistics
- Photographs, anecdotes, cartoons



ORDER FROM:

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036

(FAX) (202) 265-2384

Prepaid orders sent postage & handling free. Visa and Mastercard orders accepted. (Please give the card number and expiration date on credit card orders) We will bill for orders over $10.00.

NEW! VERSION 2

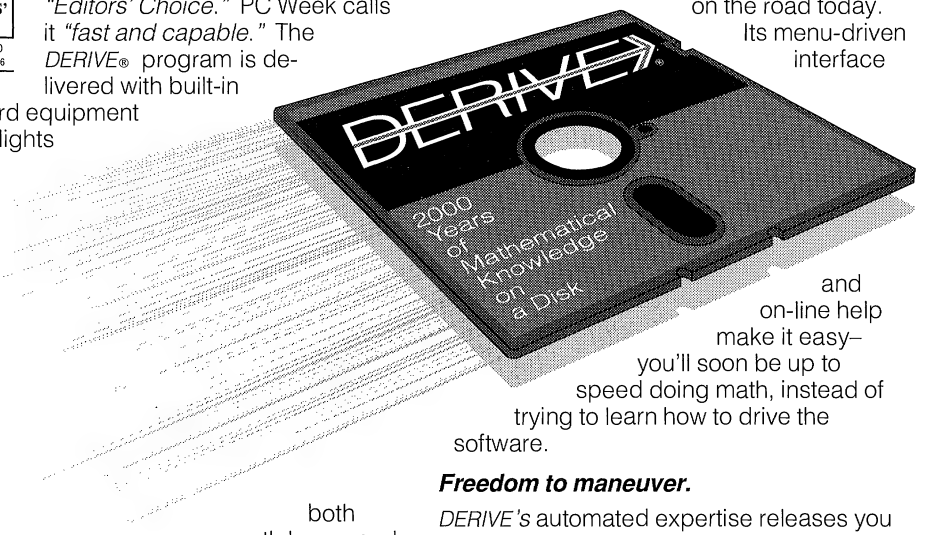
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE
May 29, 1990
Derive, Version 1 6

DERIVE® – A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU • HAWAII

Stories About Maxima and Minima

V.M. Tikhomirov,
Translated by Abe Shenitzer

We are pleased to announce the first volume in a new series of expository books translated from the Russian KVANT Library, published jointly with the American Mathematical Society.

Throughout the history of mathematics, maximum and minimum problems have played an important role. Many beautiful and important problems have appeared in a variety of branches of mathematics and physics. The greatest scientists of the past–Euclid, Archimedes, Heron, the Bernoullis and Newton, took part in seeking solutions to these concrete problems. The solutions stimulated the development of the theory, and, as a result, techniques were elaborated that made possible the solution of a tremendous variety of problems by a single method.

This book presents fifteen "stories" designed to acquaint the reader with the central concepts of the theory of maxima and minima as well as with its illustrious history.

In Part One, the author familiarizes readers with concrete problems that lead to discussion of the work of some of the greatest mathematicians of all time. Part Two introduces a method for solving maximum and minimum problems that originated with Lagrange. While the content of this method has varied constantly, its basic concept has endured for over two centuries. The final story is addressed primarily to those who teach mathematics, for it impinges on the question of how to teach. The author strives to show how the analysis of diverse facts gives rise to a general idea, how this idea is transformed, how it is enriched by new content, and how it remains the same in spite of these changes.

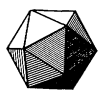
Contents

Ancient maximum and minimum problems: 1. *Why do we solve maximum and minimum problems?*; 2. *The oldest problem–Dido's problem;* 3. *Maxima and minima in nature (optics);* 4. *Maxima and minima in geometry;* 5. *Maxima and minima in algebra and in analysis;* 6. *Kepler's problem;* 7. *The brachistochrone;* 8. *Newton's aerodynamical problem;* **Methods of solution of extremal problems;** 9. *What is a function?;* 10. *What is an extremal problem?;* 11. *Extrema of functions of one variable;* 12. *Extrema of functions of many variables. Lagrange's principle;* 13. *More problem solving;* 14. *What happened later in the theory of extremal problems;* 15. *The fifteenth story, or rather, a discussion.*

187 pp., 1991, Paperbound
ISBN 0-8218-0165-1

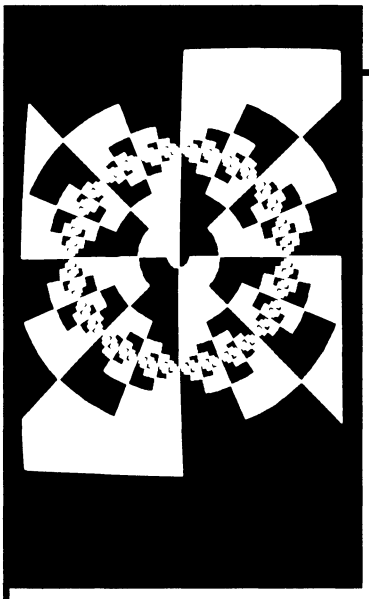
List: $23.00 MAA Member: $18.00

Catalog Number: MAXIM



ORDER FROM:
Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036
(FAX) (202) 265-2384

Prepaid orders sent postage & handling free. Visa and Mastercard orders accepted. (Please give the card number and expiration date on credit card orders) We will bill for orders over $10.00.



Fractals

Endlessly Repeated Geometrical Figures

Hans Lauwerier

Translated by Sophia Gill-Hoffstädt

Fractals are shapes in which an identical motif repeats itself on an ever diminishing scale. No longer mathematical curiosities, fractals are now a vital subject of mathematical study, practical application, and popular interest. For readers interested in graphic design, computers, and science and mathematics in general, Hans Lauwerier provides an accessible introduction to fractals that includes instructions on how to design both new and well-known examples with the help of a microcomputer.

35 color illustrations. 103 line illustrations.

Paper: $14.95 ISBN 0-691-02445-6

Cloth: $39.95 ISBN 0-691-08551-X

Not available from Princeton in the United Kingdom and Europe

D-modules and Spherical Representations

Frédéric Bien

The theory of D-modules deals with the algebraic aspects of differential equations. The book provides a general introduction to the theory of D-modules on flag varieties, and it describes spherical D-modules in terms of a cohomological formula. Using microlocalization of representations, the author derives a criterion for irreducibility. The relation between multiplicities and singularities is also discussed at length.

Mathematical Notes, 39

Paper: $22.50 ISBN 0-691-02517-7

In Japan order from United Publishers Services

With a new introduction by Thomas Banchoff

Flatland

A Romance of Many Dimensions

Edwin Abbott Abbott

Over a hundred years ago, Edwin Abbott Abbott wrote a mathematical adventure set in a world on one plane, populated by a hierarchical society of regular geometrical figures—who think and speak and have all too human emotions. Since then *Flatland* has fascinated generations of readers, becoming a perennial science-fiction favorite.

Today, we can manipulate objects in four dimensions and observe their three-dimensional slices tumbling on the computer screen. But how do we interpret these images? In his introduction to the volume, Thomas Banchoff points out that there is no better start on the problem of understanding higher-dimensional slicing phenomena than reading this classic novel of the Victorian era.

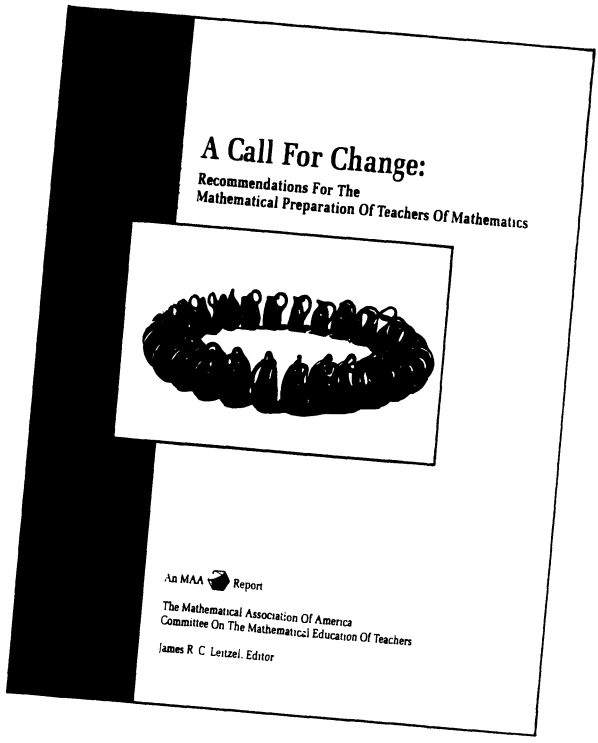
Paper: $7.95 ISBN 0-691-02525-8

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900

ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

An MAA Report on Teacher Preparation



A CALL FOR CHANGE
Recommendations for the Mathematical Preparation of Teachers of Mathematics

James R. C. Leitzel, Editor

How can we improve the teaching and learning of mathematics in our schools to better prepare our students for the future? We can begin by making some changes in the way our teachers learn and teach mathematics. A CALL FOR CHANGE, a new MAA Report, offers a set of recommendations that come from a vision of ideal mathematics teachers in classrooms of the 1990's and beyond. It describes the collegiate mathematical experiences that a teacher needs in order to meet this vision.

A CALL FOR CHANGE discusses standards common to the preparation of mathematics teachers at all levels; Elementary School Level (K–4), Middle School Level (5–8), and Secondary School Level (9–12). It shows what type of preparation teachers need to help them:

- (i) view mathematics as a system of interrelated principles,
- (ii) communicate mathematics accurately verbally, and in writing,
- (iii) understand the elements of mathematical modeling,
- (iv) understand and use calculators and computers appropriately in the teaching and learning of mathematics, and
- (v) appreciate the development of mathematics both historically and culturally.

Mathematics continues to be a dynamic changing discipline. There is new mathematics that can be exciting for young people to learn and technology provides new approaches for teachers to engage students in the teaching and learning of mathematics. The mathematical preparation of teachers must adapt to these changing realities. Order your copy now of this important set of guidelines.

60 pp., 1991, ISBN 0-88385-072-9

List: $7.00

Catalog Number CFC

Order From



Mathematical Association of America
1529 18th Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 8 October 1991

Contents

(ISSN 0002-9890)

ARTICLES

Proof of Recursive Unsolvability of Hilbert's Tenth Problem JAMES P. JONES AND YURI V. MATIJASEVIČ 689

Linear Algebra and Quantum Chemistry . . . JEROME A. GOLDSTEIN AND MEL LEVY 710

The Fifty-First William Lowell Putnam Mathematical Competition LEONARD F. KLOSINSKI, GERALD L. ALEXANDERSON, AND LOREN C. LARSON 719

NOTES

Branching Processes That Grow Faster Than Binary Splitting F. M. DEKKING 728

How Not to Repeatedly Differentiate a Reciprocal ROBERT A. LESLIE 732

Nested Radicals JONATHAN M. BORWEIN AND G. DEBARRA 735

On Infinite and Finite Covering Systems M. A. BERGER, A. FELZENBAUM, A. S. FRAENKEL, AND R. HOLZMAN 739

A Moment Convergence Theorem LAJOS TAKÁCS 742

THE TEACHING OF MATHEMATICS

Why the Method of Undetermined Coefficients Works CLAY C. ROSS, JR 747

A Freshman Calculus Proof That Real Quartics Factor JOHN ISBELL 749

On the Right-Hand Derivative of a Certain Integral Function JOHN KLIPPERT 751

Geographical Boundary Extrema HELENJANE ARMSTRONG AND ROGER HERZ-FISCHLER 752

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 754

Advanced Problems and Solutions 766

REVIEWS

How Does One Cut a Triangle? By Alexander Soifer MURRAY S. KLAMKIN 775

Concrete Mathematics: A Foundation for Computer Science. By Ronald Lewis Graham, Donald Ervin Knuth and Oren Patashnik EDWARD A. BENDER 779

TELEGRAPHIC REVIEWS 781

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $33 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

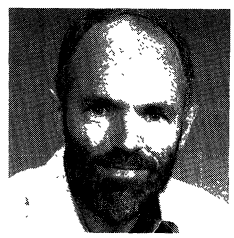
Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

Proof of Recursive Unsolvability of Hilbert's Tenth Problem

J. P. JONES, *University of Calgary, Calgary, Alberta, Canada, T2N 1N4*

Y. V. MATIJASEVIČ, *Steklov Math. Institute, Leningrad, USSR, 191011*



JAMES P. JONES received his Ph.D. in 1968 at the University of Washington, Seattle. His supervisor was R. W. Ritchie. He is presently in the Department of Mathematics of the University of Calgary, Alberta, Canada.

He has spent leaves in the Steklov Mathematical Institute of the Academy of Sciences of the U.S.S.R., Leningrad, the Tata Institute of Fundamental Research, the University of California, Berkeley and Academia Sinica, Beijing.

He has given lectures in Moscow State University, the Institute of Cybernetics in Tallinn, U.S.S.R., the Steklov Mathematical Institute in Leningrad, U.S.S.R., and in the Main Computing Center of the Academy of Sciences of the U.S.S.R. in Riga. He has also lectured at the Stefan Banach Center in Warsaw and at universities in England, France, Italy, Germany, China and Nepal.

His current research area is logic and number theory.



YURI V. MATIJASEVIČ received his Candidate of Physical and Mathematical Sciences (Ph.D.) degree in the Steklov Mathematical Institute in Leningrad, in 1970. His supervisors were N. A. Shanin and G. S. Tseitin. The degree was given for constructing a semigroup with three relations and an unsolvable word problem.

Matijosevič received his Doctor Sci. degree also in the Steklov Mathematical Institute in Leningrad, in 1973. This was for the solution of Hilbert's Tenth Problem in 1970. Then his supervisor was S. U. Maslov.

Matijasevič has given lectures in Canada, France, Hungary, Romania and the U.S. In the U.S. he has lectured in the Courant Institute, M.I.T., Boston University, Berkeley and Stanford. In Canada he has lectured in Vancouver, Calgary, London Ontario, Banff, Kingston, Quebec, Thunder Bay and Montreal. In the U.S.S.R. he has lectured in Leningrad, Moscow, Kiev, Tallinn, Kishinev and Tbilisi.

His current research area is logic and number theory.

The purpose of the present paper is to give a modern complete proof of unsolvability of the Tenth Problem of Hilbert. Our intention is to give the shortest proof known today, one which takes into account all the simplifications found since the problem was solved by the second author in [1970].

The Tenth Problem of Hilbert is the problem of solvability of diophantine equations. As originally formulated by Hilbert [1900], it was to find an algorithm to decide whether a polynomial equation in several variables, $P(x_1, x_2, \ldots, x_n) = 0$, has a solution in integers.

The problem can also be formulated in terms of existence of solutions in natural numbers (nonnegative integers). These two forms of Hilbert's Tenth problem are equivalent. An equation $P(x_1, x_2, \ldots, x_n) = 0$ has a solution in integers if and only if the product $\Pi P(\pm x_1, \pm x_2, \ldots, \pm x_n) = 0$ has a solution in natural numbers. Also, from Lagrange's Four Squares Theorem [1770], $P(x_1, x_2, \ldots, x_n) = 0$

has a solution in natural numbers iff $P(x_1^2 + y_1^2 + u_1^2 + v_1^2,\ x_2^2 + y_2^2 + u_2^2 + v_2^2, \ldots, x_n^2 + y_n^2 + u_n^2 + v_n^2) = 0$ has a solution in integers.

What will be proved here is the algorithmic unsolvability of Hilbert's Tenth Problem, i.e., the nonexistence of an algorithm, over natural numbers. This will be done in the normal way, by reducing another unsolvable problem to Hilbert's Tenth. Actually we will reduce all unsolvable problems to Hilbert's Tenth. To explain we need some definitions.

Let A be a set or relation on the natural numbers. The set A is called *recursively enumerable* (*r.e.*) if A is the exact range of a recursive function (or if A is empty). A set is called *recursive* if its characteristic function is a recursive function. A function f is said to be *recursive* if it is computable by a Turing machine or register machine (see §4).

The concept of recursively enumerable set is more general than that of recursive set. The relationship between the two concepts is also important here. A set A is recursive if and only if A and its complement A' are both r.e. sets. Thus every recursive set is recursively enumerable. However there exist r.e. sets which are non-recursive. Hence the negative solution to Hilbert's Tenth problem will follow immediately from the following theorem:

THEOREM 1. *Every r.e. relation* $A(a_1, a_2, \ldots, a_m)$ *can be represented in the form*

$$A(a_1, \ldots, a_m) \Leftrightarrow (\exists x_1, \ldots, x_n)[P(a_1, \ldots, a_m, x_1, \ldots, x_n) = 0]. \qquad (1)$$

Here $P(a_1, \ldots, a_m, x_1, \ldots, x_n)$ is a polynomial with integer coefficients which depend on A. The Theorem is that for each r.e. set A, a polynomial P exists such that (1) is satisfied for all values of $a_1, a_2, \ldots, a_m$ (the *parameters*). The variables $x_1, x_2, \ldots, x_n$, correspond to the *unknowns* in our equation. All the variables, parameters and unknowns, range over the same set, nonnegative integers.

A set or relation $A(a_1, \ldots, a_m)$ in the nonnegative integers, definable in the form (1), is called a *diophantine* set. Thus Theorem 1 asserts that every r.e. set is diophantine. The converse also holds trivially: every diophantine set is r.e. Thus Theorem 1 asserts the equality of two collections of sets.

Theorem 1 implies the nonexistence of an algorithm for Hilbert's Tenth Problem because it reduces the decision problem for every r.e. set, to some instance of Hilbert's Tenth. If Hilbert's Tenth were solvable, then every r.e. set would be recursive. However, as we mentioned, there exist r.e. sets which are nonrecursive. Taking the set A in Theorem 1 to be such a set, one sees that there can exist no algorithm for Hilbert's Tenth Problem.

This argument seems completely satisfying to us: however, for those who wish it we give an additional argument after the proof of Theorem 1 (section 5).

Theorem 1 also implies the algorithmic unsolvability of Hilbert's Tenth Problem. This is a slightly stronger conclusion than recursive unsolvability. When we come to this conclusion on the basis of Theorem 1, we are implicitly using the so-called Church–Turing Thesis. This is the statement that every computable function is Turing computable. It is called a thesis rather than a theorem because

Work supported by Natural Sciences and Engineering Research Council of Canada research grant A4525, the NSERC Program of International Scientific Exchange Awards, the Queen's–Steklov Exchange Program between Canada and the USSR and the U.S. National Academy of Sciences.

it is apparently unprovable. It is somewhat in the nature of a definition, or proposal, that we identify the intuitive concept of computability with the precise mathematical formalization, Turing computability, register machines or General Recursiveness.

Today the Church–Turing Thesis is widely accepted. If an algorithm exists to solve Hilbert's Tenth problem, then Theorem 1 implies that it would have to lie beyond the present concept of algorithm, beyond Turing machines, beyond register machines, beyond Markov algorithms and indeed beyond all known formalizations of algorithm (all of which have been proven equivalent).

Theorem 1 implies for example that an algorithm solving Hilbert's Tenth Problem would also solve the word problem for groups, the halting problem for Turing machines and all other known r.e. unsolvable problems. Needless to say, no example of such an algorithm is known.

If the reader wishes to know more about the theory of computability, he or she is directed to the books of Davis [1958], Minsky [1967], Rogers [1967], the second author's paper [1984] or the first author's [1974] MONTHLY paper.

The unsolvability of Hilbert's Tenth problem was originally proved in two steps. Step 1 was taken by Martin Davis, H. Putnam and Julia Robinson [1961] who obtained an exponential form of Theorem 1. Here P was not a polynomial but contained an exponential function, $y = a^x$. In this form Theorem 1 can be interpreted as saying that every r.e. set is exponential diophantine.

The second step in the solution of Hilbert's Tenth Problem was to show that the exponential relation itself $y = a^x$ is diophantine. This difficult last step was taken by the second author [1970]. This proved Theorem 1 and solved Hilbert's Tenth Problem in the negative.

To show that the exponential function $y = a^x$ is diophantine, divisibility properties of the sequence of Fibonacci numbers were used in [1970]. Subsequently it was seen how to do this using the sequence of solutions of the Pell equation. Proofs of Chudnovsky [1970], Davis [1971] [1973] and Kosovskii [1971] all use the sequence of solutions of Pell's equation.

Today the Pell equation gives the simplest known proof. We will also use the Pell equation here. Concerning this part of the proof, considerable credit goes to Martin Davis and Julia Robinson. Many simplifications in the present proof can be traced to their discoveries.

By way of small new improvements in the Pell equation part of the present proof, we mention (for experts), that in the main lemma on diophantine representation of the sequence of solutions of Pell's equation (Lemma 2.27), use of the Chinese remainder theorem has been eliminated. (Davis [1973] p. 246.) Those who know the subject will also find that in this Pell equation part of the proof we eliminated also the use of the axiom of existence of infinitely many solutions to the general Pell equation,

$$(\forall d)(\exists x, y)\left[d \neq \square \rightarrow x^2 - dy^2 = 1 \wedge 0 < y\right]. \tag{2}$$

This assumption (or axiom) was used in Matijasevič–Robinson [1975] (pp. 532–533). Of course the axiom is true. From a certain point of view its use doesn't matter. However, if one is concerned with formalizability of the proof in axiomatic theories, then use of this axiom is a deficiency. So we are pleased to have eliminated it.

When d is an arbitrary non-square, as is the case in (2) ($d \neq \square$), then it is difficult to prove statement (2) from a restricted set of first order axioms. For example statement (2) cannot be proved in the theory known as Bounded Arithmetic; Peano's axioms with the induction axiom

$$A(0) \wedge (\forall x)[A(x) \to A(x+1)] \to (\forall x)[A(x)], \tag{3}$$

restricted to instances where all quantified variables in the $A(x)$ are bounded by polynomials in x.

Indeed, whether the whole of the proof of unsolvability of Hilbert's Tenth problem can be formalized in Bounded Arithmetic is an open problem. J. Paris and C. Dimitracopoulos [1982] showed that this would be the case if we add one more axiom (stronger than (2)). But without this axiom the answer is not known.

This formalization problem is important because if the proof of Hilbert's Tenth Problem can be formalized in Bounded Arithmetic, then, from an idea of A. J. Wilkie, $NP = co - NP$ follows. This is why we mention that in this paper axiom (2) is used only for d of the special form, $d = a^2 - 1$. For these d axiom (2) is provable in Bounded Arithmetic. This increases the likelihood that the present proof can be formalized there.

Besides its use in formalization, the proof here can also be used as the main link in a proof that every Turing computable function is recursive. We give a method for encoding computations of arbitrary register machines which greatly simplifies the tedious arithmetization usually involved with this procedure. This method of coding we call *bit masking*. It uses a famous theorem of Lucas [1878] on congruences in binomial coefficients this method was worked out in our [1984]. Originally we made use of a classical theorem of Kummer [1852], on the power of a prime dividing a binomial coefficient. Later we discovered the simpler proof using Lucas' Theorem (see Lemma 3.10).

Nearly the whole of the present proof of Hilbert's Tenth Problem is now number theory. In fact it is classical number theory. Section 2, on the sequence of solutions of Pell's equation, belongs to the Lucas–Lehmer theory of recurrent sequences. In the terminology of Lehmer [1930], this is the study of the Lucas sequences U_n and V_n where $P = 2a$, $Q = 1$, $D = 4(a^2 - 1)$ and $R = 4a^2$. With these values of the parameters, U_n and V_n correspond to the sequence of solutions of Pell's equation, $x^2 - dy^2 = 1$ with $d = a^2 - 1$. Here $U_n = Y_a(n)$ and $V_n = 2 \cdot X_a(n)$ where $X_a(n)$ and $Y_a(n)$ are the sequence of solutions of $x^2 - (a^2 - 1)y^2 = 1$, (see §2).

This theory is very old. Though the multiplication by 2 makes a small difference, it is not difficult to interpret $Y_a(n)$ and $X_a(n)$ in terms of U_n and V_n and so to see that many of the theorems go back to Lucas [1878] and Lehmer [1930]. Some modern theorems are due to Julia Robinson [1952] [1969]. In this connection it is interesting that while Lucas preferred the functions U_n and V_n, Julia Robinson preferred the sequences $Y_a(n)$ and $X_a(n)$.

§1 Diophantine Sets. A diophantine equation is a polynomial equation, $P(a_1, \ldots, a_m, x_1, \ldots, x_n) = 0$, in several variables, with integer coefficients. The variables are divided into *parameters* $a_1, a_2, \ldots, a_m$, and *unknowns*, $x_1, x_2, \ldots, x_n$. All the variables, parameters and unknowns, range over the set of nonnegative integers, $0, 1, 2, \ldots$.

In the classical theory of diophantine equations one begins with an equation and asks for values of the parameters for which there exists a solution. In this proof we turn the usual procedure around. We start with the solution and search for the equation. We begin with a relation $A(a_1, a_2, \ldots, a_m)$ and we look for a polynomial $P(a_1, \ldots, a_m, x_1, \ldots, x_n)$ defining it in the sense of (1).

DEFINITION 1.1. A relation $A(a_1, a_2, \ldots, a_m)$ is *diophantine* if there exists a polynomial $P(a_1, \ldots, a_m, x_1, \ldots, x_n)$ such that for all values of $a_1, \ldots, a_m$, (the parameters)

$$A(a_1, \ldots, a_m) \Leftrightarrow (\exists x_1, x_2, \ldots, x_n)[P(a_1, \ldots, a_m, x_1, \ldots, x_n) = 0]. \quad (1.1)$$

This definition which is for relations, will also do for functions. A *function* will be said to be diophantine if its graph is diophantine. Below are examples of diophantine functions and relations. Most will be used in the proof. These examples include the elementary relations of *order* $\leqslant$, *divisibility* $a|b$, and *congruence*, $a \equiv b \pmod{c}$.

$$a \leqslant b \Leftrightarrow (\exists x)[a + x = b], \quad (1.2)$$

$$a|b \Leftrightarrow (\exists x)[ax = b], \quad (1.3)$$

$$a \equiv b \pmod{c} \Leftrightarrow (\exists x)[a = b + cx \text{ or } a = b - cx]. \quad (1.4)$$

Disjunctions and conjunctions as occur in connection with (1.4) can be dealt with using

$$A = 0 \text{ or } B = 0 \Leftrightarrow A \cdot B = 0, \qquad A = 0 \,\&\, B = 0 \Leftrightarrow A^2 + B^2 = 0. \quad (1.5)$$

In the case of conjunctions, it is necessary to rename variables which occur in both A and B, for example $\exists x A(x) = 0 \wedge \exists x B(x) = 0$ is equivalent to $\exists x \exists y[A(x) = 0 \wedge B(y) = 0]$.

From (1.5) it follows that conjunctions and disjunctions of Diophantine relations (but not negations) are diophantine. Proceeding in this way, we can prove that a great many relations are diophantine. A good is the relation of *coprimality* (a is *relatively prime* to b, $(a, b) = 1$) written in this paper as $a \perp b$. This relation can be seen to be diophantine using (1.5) together with

$$a \perp b \Leftrightarrow (\exists x, y)[ax - by = 1 \text{ or } ax - by = -1]. \quad (1.6)$$

Other examples of diophantine functions are the *remainder function*, $r = rem(a, b)$ (r is the remainder after a is divided by b), and the *quotient function*, $q = quo(a, b)$, meaning q is the quotient when a is divided by b. (This is the same as the integer part function, $q = \lfloor a/b \rfloor$.) These functions can be seen to be diophantine from

$$r = rem(a, b) \leftrightarrow r \equiv a \pmod{b} \text{ and } r < b, \quad (1.7)$$

$$q = quo(a, b) \leftrightarrow 0 \leqslant a - qb < b. \quad (1.8)$$

Proceeding in this way, defining new relations from old, using, often (1.5) we will show that a great many relations are diophantine. The most important tool in this process will be the sequence of solutions of the Pell equation.

§2. The Pell equation. A general Pell equation is an equation of the form

$$x^2 - dy^2 = 1 \qquad (d \neq \square). \quad (2.1)$$

where d is a constant and x and y are unknowns. (A hyperbola in the x, y plane.)

When $d = \square$ (d is a square), then the Pell equation (2.1) has only the trivial solution, $(1, 0)$. So one always assumes that $d \neq \square$ (d is not a square).

Because $x^2 - dy^2$ factors over the reals, $x^2 - dy^2 = (x + \sqrt{d}\,y)(x - \sqrt{d}\,y)$, it is natural to work in the integral domain $\mathbb{Z}[\sqrt{d}\,]$ consisting of real numbers of the form $\alpha = x + y\sqrt{d}$, $x, y \in \mathbb{Z}$. Since $\sqrt{d}$ is irrational, the integers x and y are unique. They are called the *components* of α.

When $\alpha = x + y\sqrt{d}$, the number $\bar{\alpha} = x - y\sqrt{d}$ is known as the *conjugate* of α. The number $N(\alpha) = \alpha\bar{\alpha} = x^2 - dy^2$ is called the *norm* of α. Since $N(\alpha) = x^2 - dy^2$, the integer solutions to the Pell equation are just the components x, y of reals α such that $N(\alpha) = 1$.

$N(\alpha) = 1$ and $N(\alpha) = \alpha \cdot \bar{\alpha}$ imply $\alpha^{-1} = \bar{\alpha}$, the conjugate of α is equal to the inverse. Also $\alpha = x + \sqrt{d} \cdot y$ and $1 \leqslant \alpha$ implies $1 \leqslant x$ and $0 \leqslant y$. To see this observe that from $0 < \alpha^{-1} \leqslant 1$, $\alpha + \bar{\alpha} = 2x$ and $\alpha - \bar{\alpha} = 2y\sqrt{d}$ we have $\frac{1}{2} \leqslant x$ and $0 \leqslant y$. Then from $x^2 = 1 + dy^2$ we have $1 \leqslant x$. Hence the inequality $1 \leqslant \alpha$ implies that the components x and y of α are nonnegative.

From $\overline{\alpha\beta} = \bar{\alpha} \cdot \bar{\beta}$ we have $N(\alpha\beta) = \alpha\beta \cdot \overline{\alpha\beta} = \alpha \cdot \beta \cdot \bar{\alpha} \cdot \bar{\beta} = \alpha \cdot \bar{\alpha} \cdot \beta \cdot \bar{\beta} = N(\alpha) \cdot N(\beta)$. Thus if $N(\alpha) = 1$ and $N(\beta) = 1$, then $N(\alpha \cdot \beta) = 1$. So the product of two reals representing solutions of the Pell equation represents again a solution. We also have $N(\bar{\alpha}') = N(\alpha)$.

Now let $\alpha = x_1 + y_1\sqrt{d}$ and $\beta = x_2 + y_2\sqrt{d}$ be two solutions of the Pell equation, with x_1, y_1, x_2, y_2 integers, $N(\alpha) = 1$ and $N(\beta) = 1$. Suppose $1 \leqslant \alpha$ and $1 \leqslant \beta$ hold. As we have seen, x_1, y_1, x_2, y_2 must then be nonnegative. So from the equations $x_1^2 = dy_1^2 + 1$ and $x_2^2 = dy_2^2 + 1$, it follows that $x_1 < x_2$ iff $y_1 < y_2$. Thus $1 \leqslant \alpha < \beta$ holds iff both $1 \leqslant x_1 < x_2$ and $0 \leqslant y_1 < y_2$ hold. So the set of reals α for which $N(\alpha) = 1$ and $1 < \alpha$ hold, is a well ordered set. If this set is nonempty, if there exists a nontrivial solution to the Pell equation, then there must exist a least real α such that $1 < \alpha$ and $N(\alpha) = 1$. This real is called the *generator*. Its components are called the *fundamental solution*.

The powers of the generator α generate all solutions to (2.1). To see this, observe first that $N(\alpha) = 1$ implies $N(\alpha^n) = N(\alpha)^n = 1$, so that powers of α represent solutions. To see that these powers of α give all solutions of (2.1), observe that if $N(\beta) = 1$ and $1 \leqslant \beta$, then $\exists n$ such that $\alpha^n \leqslant \beta < \alpha^{n+1}$. From $1 \leqslant \beta \cdot \alpha^{-n} < \alpha$ and $N(\beta\alpha^{-n}) = N(\beta)N(\alpha^{-1})^n = N(\beta)N(\bar{\alpha})^n = N(\beta)N(\alpha)^n = 1$, it then follows that $\beta \cdot \alpha^{-n} = 1$. Hence $\beta = \alpha^n$. Therefore β can be obtained as a power of α.

It is not entirely trivial to show that for arbitrary $d \neq \square$ there always exists a nontrivial solution to the Pell equation (2.1). But when d is a nonsquare of the special form $d = a^2 - 1$, then it is easy to prove this and hence the existence of infinitely many solutions. The special Pell equation with $d = \alpha^2 - 1$,

$$x^2 - (a^2 - 1)y^2 = 1, \tag{2.2}$$

has the fundamental solution $(x, y) = (a, 1)$. So the generator is $a + \sqrt{a^2 - 1}$. By the preceding, the powers of this real then give all solutions,

$$X_a(n) + Y_a(n) \cdot \sqrt{a^2 - 1} = \left(a + \sqrt{a^2 - 1}\right)^n. \tag{2.3}$$

Let $x = X_a(n)$ and $y = Y_a(n)$ denote this sequence of solutions to (2.2), defined by (2.3). Then $X_a(n)$ and $Y_a(n)$ are strictly increasing functions of n. This is evident from (2.9) below.

Taking the conjugate of both sides of (2.3) one finds that the sequences $X_a(n)$ and $Y_a(n)$ are also definable from the conjugate of the generator, $a - \sqrt{a^2 - 1}$,

$$X_a(n) - Y_a(n)\sqrt{a^2 - 1} = \left(a - \sqrt{a^2 - 1}\right)^n. \tag{2.4}$$

As we have mentioned, $\bar{\alpha} = \alpha^{-1}$, the conjugate of the generator is equal to the inverse,

$$a - \sqrt{a^2 - 1} = \left(a + \sqrt{a^2 - 1}\right)^{-1}. \tag{2.5}$$

This implies that most of the identities which hold for the sequence of solutions of the Pell equation also hold for negative values of parameters. For example from the identities $(a + \sqrt{d})^{n+m} = (a + \sqrt{d})^n(a + \sqrt{d})^m$, $(a + \sqrt{d})^{n-m} = (a + \sqrt{d})^n(a + \sqrt{d})^{-m}$, using (2.3), (2.4) and (2.5) one obtains the identity

$$X_a(n \pm m) + Y_a(n \pm m)\sqrt{d} = \left(X_a(n) + Y_a(n)\sqrt{d}\right)\left(X_a(m) \pm Y_a(m)\sqrt{d}\right). \tag{2.6}$$

Taking the rational and irrational parts of (2.6), one obtains the Addition Equations (Lucas). Here the $\pm$ signs correspond.

$$X_a(n \pm m) = X_a(n)X_a(m) \pm dY_a(n)Y_a(m), \tag{2.7}$$

$$Y_a(n \pm m) = Y_a(n)X_a(m) \pm X_a(n)Y_a(m). \tag{2.8}$$

These hold only for $a \geqslant 2$ but if we define $X_1(n) = 1$ and $Y_1(n) = n$, then (2.8) will hold for $a = 1$. Putting $m = 1$ in (2.7) and (2.8), we see that also, as a special case of (2.7), (2.8)

$$X_a(n + 1) = aX_a(n) + dY_a(n), \qquad Y_a(n + 1) = aY_a(n) + X_a(n), \tag{2.9}$$

$$X_a(n - 1) = aX_a(n) - dY_a(n), \qquad Y_a(n - 1) = aY_a(n) - X_a(n). \tag{2.10}$$

Adding pairs of equations (2.9) and (2.10), one obtains representations of X and Y as *Lucas Sequences*, that is, sequences satisfying a second order linear recurrence.

$$X_a(0) = 1, \quad X_a(1) = a, \quad X_a(n + 1) = 2aX_a(n) - X_a(n - 1). \tag{2.11}$$

$$Y_a(0) = 0, \quad Y_a(1) = 1, \quad Y_a(n + 1) = 2aY_a(n) - Y_a(n - 1). \tag{2.12}$$

From the addition equations, (2.7) and (2.8), together with defining equation (2.2), one can derive the *Double Angle Formulas* (Lucas)

$$X_a(2n) = 2X_a(n)^2 - 1, \tag{2.13}$$

$$Y_a(2n) = 2X_a(n)Y_a(n). \tag{2.14}$$

Generally we are interested in the X_a and Y_a sequences only for $a \geqslant 2$. However as we mentioned, when $a = 1$ they can be defined in a natural way satisfying (2.8), (2.11) and (2.12). We can define $X_1(n) = 1$ and $Y_1(n) = n$.

For fixed n, the function $Y_a(n)$ is a polynomial in a. This can be seen from (2.12). The degree is $n - 1$. From this, that $Y_a(n)$ is a polynomial in a, one obtains

Congruence Rule, $$Y_a(n) \equiv Y_b(n) \pmod{a - b}. \tag{2.15}$$

The Congruence Rule holds for $a, b \geqslant 1$. Putting $b = 1$ in (2.15) and using $Y_1(n) = n$ we obtain the special congruence rule of Julia Robinson [1952]:

$$Y_a(n) \equiv n \pmod{a - 1}. \tag{2.16}$$

The Lucas equations (2.11), (2.12) also enable one to derive bounds on the size of the X_a and Y_a sequences. For example, it is easy to derive the following bound on $Y_a(n)$:

$$(2a-1)^n \leqslant Y_a(n+1) < (2a)^n. \tag{2.17}$$

This inequality, which holds for $1 \leqslant a$, $1 < n$, shows that the $Y_a(n)$ sequence grows exponentially in n. These sequences can also be more directly related to exponentiation. The following congruence was obtained by Julia Robinson [1952]:

$$X_a(n) - (a-k)Y_a(n) \equiv k^n \pmod{2ak - k^2 - 1}. \tag{2.18}$$

Proof. We will show that the congruence holds for all $n \geqslant 0$ and all $k \geqslant 0$. It is easy to check for $n = 0$ and $n = 1$. So proceeding by induction on n, using recurrence equations (2.11) and (2.12), we obtain

$$\begin{aligned} X_a(n+1) - (a-k)Y_a(n+1) &= 2a \cdot X_a(n) - X_a(n-1) \\ &\quad -(a-k)\left[2a \cdot Y_a(n) - Y_a(n-1)\right] \\ &= 2a\left[X_a(n) - (a-k)Y_a(n)\right] \\ &\quad -\left[X_a(n-1) - (a-k)Y_a(n-1)\right] \\ &\equiv 2ak^n - k^{n-1} = k^{n-1}(2ak-1) \\ &= k^{n-1}(2ak - k^2 - 1 + k^2) \equiv k^{n-1}(0 + k^2) \\ &= k^{n-1}k^2 = k^{n+1} \pmod{2ak - k^2 - 1}. \end{aligned}$$

Next we prove a divisibility property which we will need in the proof of (2.20).

$$n|m \leftrightarrow Y_a(n)|Y_a(m). \tag{2.19}$$

Proof. From the Addition Equation (2.8) we have (dropping the subscript a), $Y(k \pm n) = Y(k)X(n) \pm X(k)Y(n) \equiv Y(k)X(n) \pmod{Y(n)}$. But $Y(n) \perp X(n)$ ($Y(n)$ and $X(n)$ are relatively prime), by (2.2). So $Y(n)|Y(k \pm n)$ iff $Y(n)|Y(k)$. Now let $m = ni + r$ where $0 \leqslant r < n$. Then $0 \leqslant Y(r) < Y(n)$. Also $Y(n)|Y(m)$ iff $Y(n)|Y(ni + r)$ iff $Y(n)|Y(r)$. Hence $Y(n)|Y(m)$ iff $r = 0$, i.e. $Y(n)|Y(m)$ iff $n|m$.

FIRST STEP DOWN LEMMA 2.20. $Y_a^2(n)|Y_a(m) \leftrightarrow n \cdot Y_a(n)|m$. $(1 \leqslant a)$.

Proof. By using identity (2.3), twice, it is easy to prove that for any j

$$X_a(nj) + Y_a(nj)\sqrt{d} = \left(X_a(n) + Y_a(n)\sqrt{d}\right)^j. \tag{2.21}$$

Now expand the right side of (2.21) and take the irrational part to get

$$Y_a(nj) = \sum_{i \text{ odd}}^{j} \binom{j}{i} X_a(n)^{j-i} Y_a(n)^i \left(\sqrt{d}\right)^{i-1}. \tag{2.22}$$

Hence

$$Y_a(nj) \equiv jX_a(n)^{j-1}Y_a(n) \left(\mathrm{mod}\, Y_a(n)^3\right). \tag{2.23}$$

For the proof of 2.20 in the $\Rightarrow$ direction, suppose $Y(n)^2|Y(m)$. Then by (2.19) $m = nj$ for some j. Let this be the j in (2.23). Since $X(n) \perp Y(n)$, (2.23) implies $Y(n)^2|jY(n)$. Hence $Y(n)|j$, so that $n \cdot Y(n)|m$. For the converse suppose $n \cdot Y(n)|m$. Let $j = Y(n)$. Then (2.23) $\Rightarrow Y(n)^2|Y(n \cdot Y(n))$. So by (2.19), $Y(n)^2|Y(m)$.

LEMMA 2.24. *For* $2 \leqslant a$ *and* $1 \leqslant n$, $Y_a(n-1) + Y_a(n) < X_a(n)$.

Proof. Replace n by $n-1$ in (2.9) to obtain $a \cdot Y(n-1) + X(n-1) = Y(n)$. From $2 \leqslant a$, we get $2 \cdot Y(n-1) \leqslant a \cdot Y(n-1) < a \cdot Y(n-1) + X(n-1) = Y(n)$. Hence $Y(n-1) < Y(n) - Y(n-1)$. Add $Y(n)$ to both sides to obtain $Y(n-1) + Y(n) < 2 \cdot Y(n) - Y(n-1) \leqslant a \cdot Y(n) - Y(n-1) = X(n)$, by (2.10).

LEMMA 2.25. $Y_a(4ni \pm m) \equiv Y_a(m)$, $Y_a(4ni + 2n \pm m) \equiv \mp Y_a(m) \pmod{X_a(n)}$.

Proof. From (2.14) $Y(2n) \equiv 0 \pmod{X(n)}$. From (2.13) $X(2n) \equiv -1 \pmod{X(n)}$. So from (2.8) $Y(2n \pm m) \equiv Y(2n) \cdot X(m) \pm X(2n) \cdot Y(m) \equiv \mp Y(m) \pmod{X(n)}$, we have $Y(2n \pm m) \equiv \mp Y(m) \pmod{X(n)}$. Using this $Y(4n \pm m) = Y(2n + 2n \pm m) \equiv -Y(2n \pm m) \equiv \pm Y(m) \pmod{X(n)}$. So also $Y(4n \pm m) \equiv \pm Y(m) \pmod{X(n)}$. Here the signs $\pm$ correspond. In the next lemma they do not.

SECOND STEP DOWN LEMMA 2.26. $Y_a(k) \equiv \pm Y_a(m) \pmod{X_a(n)} \leftrightarrow k \equiv \pm m \pmod{2n}$.

Proof. We assume as usual $2 \leqslant a$, $1 \leqslant n$. For the proof in the $\Leftarrow$ direction, suppose that $k = 2nj \pm m$. When $j = 2i$, we have $Y(k) = Y(4ni \pm m) \equiv \pm Y(m) \pmod{X(n)}$. In the case that $j = 2i + 1$, we have $Y(k) = Y(4ni + 2n \pm m) \equiv \mp Y(m) \pmod{X(n)}$. So $Y(k) \equiv \pm Y(m) \pmod{X(n)}$.

For the proof in the $\Rightarrow$ direction, assume that $Y(k) \equiv \pm Y(m) \pmod{X(n)}$. Choose k' such that $0 \leqslant k' \leqslant n$ and $k \equiv \pm k' \pmod{2n}$. Choose m' such that $0 \leqslant m' \leqslant n$ and $m \equiv \pm m' \pmod{2n}$. Then from the assumption and Lemma 2.26 in the direction already proven, $\Leftarrow$, we have $Y(k') \equiv \pm Y(m') \pmod{X(n)}$. Hence it follows that $X(n) | Y(k') \pm Y(m')$. Therefore $k' = m'$, because if we suppose $k' \neq m'$, then we would have $0 < |Y(k') \pm Y(m')| \leqslant |Y(k') + Y(m')| \leqslant Y(n-1) + Y(n) < X(n)$, by Lemma 2.24. From $k' = m'$ we deduce $k \equiv \pm m \pmod{2n}$.

LEMMA 2.27. *For* $A > 1$. *In order that* $C = Y_A(B)$, *it is necessary and sufficient that there exist natural numbers* D, E, F, G, H, I *and* i *such that*

(1) $D^2 - (A^2 - 1)C^2 = 1$, (4) $E = (i+1)2C^2$, (7) $H \equiv C \pmod{F}$,

(2) $F^2 - (A^2 - 1)E^2 = 1$, (5) $G \equiv A \pmod{F}$, (8) $H \equiv B \pmod{2C}$,

(3) $I^2 - (G^2 - 1)H^2 = 1$, (6) $G \equiv 1 \pmod{2C}$, (9) $B \leqslant C$.

Proof of Sufficiency: Suppose there exist D, E, F, G, H, I, i satisfying (1)–(9). Then equations (1)–(3), Pell equations, imply the existence of numbers p, q, and r such that

$$D = X_A(p), \quad C = Y_A(p), \quad F = X_A(q),$$
$$E = Y_A(q), \quad I = X_G(r), \quad H = Y_G(r).$$

We also have $0 \leqslant p \leqslant C$ and $0 \leqslant B \leqslant C$. Hence the idea is to show $B = p$, by proving $B \equiv r \equiv \pm p \pmod{2C}$. We can suppose $0 < C$. Using the First Step Down Lemma, together with (4), we have

$$C^2 | E \Rightarrow Y_A^2(p) | Y_A(q) \Rightarrow Y_A(p) | q \Rightarrow C | q.$$

To show that $B \equiv r \pmod{2C}$ we use (6), (8) and the Congruence Rule (2.16) to get

$$B \equiv H = Y_G(r) \equiv Y_1(r) = r \pmod{2C} \Rightarrow B \equiv r \pmod{2C}. \quad (2.28)$$

By the Second Step Down Lemma, the Congruence Rule (2.15), by (5) and by (7),

$$Y_A(r) \equiv Y_G(r) = H \equiv C = Y_A(p) \pmod{X_A(q)} \Rightarrow r \equiv \pm p \pmod{2q}.$$

But $C|q$. Hence $r \equiv \pm p \pmod{2C}$. This plus (2.28) implies $B \equiv \pm p \pmod{2C}$.

Proof of Necessity: Suppose $C = Y_A(B)$. Let $D = X_A(B)$. Then (1) and (9) hold. Put $q = B \cdot Y_A(B)$, $F = X_A(2q)$ and $E = Y_A(2q)$. Then (2) holds. Let $m = B \cdot Y_A(B)$ in the First Step Down Lemma (2.20). Then the First Step Down Lemma says

$$Y_A(B)^2 | Y_A(B \cdot Y_A(B)). \quad (2.29)$$

Hence $C^2|Y_A(q)$. The Double Angle Formula (2.14) says $2X_A(q) \cdot Y_A(q)|Y_A(2q)$. Hence $2C^2|E$. Therefore (4) can be satisfied. Put $G = A + F^2(F^2 - A)$. Then (5) holds. Also (2) and (4) together imply that $F^2 \equiv 1 \pmod{2C}$. Then $G = A + F^2(F^2 - A)$ implies (6). Put $I = X_G(B)$ and $H = Y_G(B)$. Then (3) holds. From (2.16) $H = Y_G(B) \equiv B \pmod{G - 1}$. So by (6) $H \equiv B \pmod{2C}$. Therefore (8) holds. From (2.15) we have $H = Y_G(B) \equiv Y_A(B) = C \pmod{G - A}$. This together with (5) implies $H \equiv C \pmod{F}$. Hence (7) holds.

§3. Exponential and Binomial Coefficient. By Lemma 2.27, the 3-place relation $y = Y_a(n)$ is diophantine. Hence we may use it in showing that other relations are diophantine. We can also use the 3-place relation $x = X_a(n)$ for this purpose. It is also diophantine. (That $x = X_a(n)$ is diophantine follows from Lemma 2.27 and equation (2.2).) We will use both $y = Y_a(n)$ and $x = X_a(n)$ in defining the exponential relation.

LEMMA 3.1. *Suppose* $1 \leqslant n$, $2 \leqslant k$. *For a sufficiently large,* $a \geqslant Y_k(n + 1)$

$$k^n = rem\big(X_a(n) - (a - k)Y_a(n),\ 2ak - k^2 - 1\big).$$

Proof. From (2.17) we have $k \leqslant k^n < (2k - 1)^n \leqslant Y_k(n + 1) \leqslant a$. From this it follows that $k + 1 \leqslant a$, and hence $a < a \cdot k < a \cdot k + k - 1 = ak + (k + 1)k - k^2 - 1 \leqslant ak + ak - k^2 - 1 = 2ak - k^2 - 1$. Therefore $k^n < 2ak - k^2 - 1$, k^n is smaller than the modulus. But by congruence (2.18), $k^n \equiv X_a(n) - (a - k)Y_a(n) \pmod{2ak - k^2 - 1}$. Since k^n is less than the modulus, it is the remainder.

Lemmas 2.27, 3.1 and (1.7) imply that the exponential relation, $m = k^n$ is diophantine. By Lemma 3.1, $m = k^n$ if and only if there exists an a such that

$$\begin{aligned} &m \equiv X_a(n) - (a - k) \cdot Y_a(n) \pmod{2ak - k^2 - 1}, \\ &m < a \qquad a \geqslant Y_k(n + 1). \end{aligned} \quad (3.1)$$

We show next that the binomial coefficient is diophantine. This relation was first proved exponential diophantine by Julia Robinson [1952]. The basic idea of the proof is that the binomial coefficients are just the digits in the base u expansion of $(1 + u)^n$, when u is large enough.

LEMMA 3.2. *For* $0 \leqslant k$, n *and* $u > 2^n$,

$$\binom{n}{k} = rem\left(\left\lfloor \frac{(u+1)^n}{u^k} \right\rfloor, u\right).$$

Proof. Expand $(u+1)^n$ by the Binomial Theorem and divide by u^k to get,

$$\frac{(u+1)^n}{u^k} = \sum_{i=k+1}^{n} \binom{n}{i} u^{i-k} + \binom{n}{k} + \sum_{i=0}^{k-1} \binom{n}{i} u^{i-k}. \tag{3.2}$$

From (3.2) together with the inequality $u^{i-k} \leqslant 1/u$, which holds for $i \leqslant k-1$, we get

$$\sum_{i=0}^{k-1} \binom{n}{i} u^{i-k} \leqslant \sum_{i=0}^{k-1} \binom{n}{i} \frac{1}{u} \leqslant \frac{1}{u} \sum_{i=0}^{n} \binom{n}{i} = \frac{2^n}{u} < 1. \tag{3.3}$$

Hence

$$\left\lfloor \frac{(u+1)^n}{u^k} \right\rfloor \equiv \binom{n}{k} \pmod{u}.$$

So Lemma 3.2 follows from $\binom{n}{k} \leqslant 2^n < u$.

Lemmas 2.27, 3.1, 3.2, 1.7 and 1.8 imply that the binomial coefficient is diophantine. We have $m = \binom{n}{k}$ iff there exist u, x, and y, such that

$$(u+1)^n = y \cdot u^{k+1} + m \cdot u^k + x, \qquad 2^n < u, \qquad x < u^k \qquad \text{and} \qquad m < u. \tag{3.4}$$

We now use this to prove that the following relation $\preccurlyeq$ (*bit masking*) is diophantine.

DEFINITION 3.5. *Let* r *and* s *be natural numbers written in binary,* (*base* 2). *Then* $r \preccurlyeq s$ *means each binary digit of* r *is less than or equal to the corresponding binary digit of* s.

The masking relation $\preccurlyeq$ is closely related to the operation of taking the logical *and* of two numbers. One can define $\preccurlyeq$ from logical and by

$$a \preccurlyeq b \Leftrightarrow a \wedge b = a. \tag{3.6}$$

Logical and is definable from $\preccurlyeq$ by

$$a \wedge b = c \Leftrightarrow c \preccurlyeq a \qquad \text{and} \qquad a \preccurlyeq a + b - c. \tag{3.7}$$

Another property of the masking relation is that $r \preccurlyeq s$ implies $r \leqslant s$. The masking relation also satisfies the axioms for a partial ordering, reflexivity, anti-symmetry and transitivity.

The masking relation $\preccurlyeq$ is also diophantine. For the proof we will use the Theorem of E. Lucas [1878] on binomial coefficients modulo a prime.

THEOREM 3.8 (Lucas [1878]). *If* p *is prime,* r_i, s_i *the base* p *digits of* r *and* s, *then*

$$\binom{s}{r} = \binom{s_n}{r_n} \cdots \binom{s_1}{r_1}\binom{s_0}{r_0} \pmod{p}. \tag{3.8}$$

Proof. It is easy to see, by induction on n, that this general form of Lucas Theorem, Theorem 3.8, is implied by the following special case (H. Anton [1869], approximately $n = 1$), of Lucas' Theorem:

LEMMA 3.9. *Let p be a prime. Suppose $0 \leqslant b < p$ and $0 \leqslant d < p$. Then*

$$\binom{a \cdot p + b}{c \cdot p + d} \equiv \binom{a}{c} \cdot \binom{b}{d} \pmod{p}. \tag{3.9}$$

Proof. Working in the polynomial ring $\mathbb{Z}_p[x]$, consider the coefficient of the term x^{cp+d} in the polynomial $(1 + x)^{ap+b}$. From the equality $(x + y)^p = x^p + y^p$ we get

$$\sum_{n=0}^{ap+b} \binom{ap + b}{i} x^n = (1 + x)^{a \cdot p + b} = (1 + x)^{pa}(1 + x)^b = (1 + x^p)^a \cdot (1 + x)^b$$

$$= \left(\sum_{j=0}^{a} \binom{a}{j} x^{jp} \right) \left(\sum_{i=0}^{b} \binom{b}{i} x^i \right) = \sum_{j=0}^{a} \sum_{i=0}^{b} \binom{a}{j} \binom{b}{i} x^{jp+i}.$$

The coefficient of x^{cp+d} must be the same in both polynomials. However because of the inequalities $0 \leqslant i \leqslant b < p$ and $0 \leqslant d < p$, the equation $jp + i = cp + d$ can hold only when $i = d$ and $j = c$. Therefore, the coefficient of the term x^{cp+d} in $(1 + x)^{ap+b}$ must therefore be $\binom{a}{c}\binom{b}{d}$. Hence $\binom{ap + b}{cp + d} = \binom{a}{c} \cdot \binom{b}{d}$ holds in $\mathbb{Z}_p[x]$.

Now we can prove that the masking relation $\preccurlyeq$ is diophantine. This will follow from

LEMMA 3.10.

$$r \preccurlyeq s \Leftrightarrow \binom{s}{r} \equiv 1 \pmod{2}.$$

Proof. Put $p = 2$ in Theorem 3.8. Then 3.10 follows from the trivial relations

$$\binom{1}{0} = 1, \qquad \binom{1}{1} = 1, \qquad \binom{0}{0} = 1, \qquad \binom{0}{1} = 0.$$

Lemma 3.10 is proved. From this Lemma together with Lemma 3.10, 3.2, (1.7) and (1.8), it follows that $\preccurlyeq$ is a diophantine relation.

§4. Arithmetization of register machines. We now have enough diophantine relations to prove that every recursively enumerable set is diophantine. Recall the definition of r.e. set: A set A is r.e. if and only if A is empty or the range of a recursive function. For the definition of recursive function, we will use the register machine, a model of computation equivalent to the Turing machine. A *register machine* is a "machine" with a finite program and a finite number of separately addressable registers, $R1, R2, \ldots, Rr$. The registers are assumed unbounded, each register can contain an arbitrarily large nonnegative integer. A subset of the registers, say $R_1, R_2, \ldots, R_k$, $(k < r)$, is designated as *input* registers and a subset, say $R_1, R_2, \ldots, R_m$, $(m < r)$, as *output* registers. This is to handle functions of k variables whose values may be m-tuples. For a function of one variable of course $k = 1$. Normally also $m = 1$. Usually R_1 is thought of as the input–output register.

A register machine is actually a program, a list of commands written on lines labelled $L1, L2, \ldots, Ll$. The register machine's commands are normally executed

in sequence; however, the register machine may execute commands which tell it to transfer to a different location and begin execution there.

For the computation of all recursive functions, it is sufficient to assume that the register machine is capable of adding or subtracting 1 from a register and of transferring on zero.

The subtraction command can cause a problem if one attempts to subtract 1 from an already zero register. For this reason Minsky [1967] permitted the subtraction command to occur only after a test for zero. His subtraction command therefore required two lines

$$Li \qquad \text{IF } Rj = 0, \text{ GOTO } Lk, \tag{4.1}$$

$$Li + 1 \quad \text{ELSE } Rj \leftarrow R_j - 1. \tag{4.2}$$

It is sufficient to assume that the program is written in such a way that subtraction from a zero register never occurs. We will assume that our machine has the following one line commands:

	COMMAND	INTERPRETATION	
Li	GOTO Lk	*Transfer to line* Lk.	(4.3)
Li	IF $0 < Rj$ GOTO Lk	*Conditional transfer to line* Lk.	(4.4)
Li	$Rj \leftarrow Rj + 1$,	*Increment register* Rj.	(4.5)
Li	$Rj \leftarrow Rj - 1$,	*Decrement register* Rj.	(4.6)

In fact because it shortens our example and does not complicate the proof, we will allow several commands of type (4.5) and (4.6) to be written on one line, in parallel, when they refer to different registers. Below is an example of a register machine.

EXAMPLE 1. *A register machine which computes the* x^{th} *Fibonacci number,* F_x.

$L1$ IF $R1 = 0$, GOTO $L20$
$L2$ $R2 \leftarrow R2 + 1$, $R3 \leftarrow R3 + 1$
$L3$ $R1 \leftarrow R1 - 1$
$L4$ IF $R1 = 0$, GOTO $L16$
$L5$ $R1 \leftarrow R1 - 1$
$L6$ $R4 \leftarrow R4 + 1$, $R5 \leftarrow R5 + 1$
$L7$ $R3 \leftarrow R3 - 1$
$L8$ IF $0 < R3$, GOTO $L6$
$L9$ $R4 \leftarrow R4 + 1$, $R2 \leftarrow R2 - 1$
$L10$ IF $0 < R2$, GOTO $L9$
$L11$ $R3 \leftarrow R3 + 1$, $R4 \leftarrow R4 - 1$
$L12$ IF $0 < R4$, GOTO $L11$
$L13$ $R2 \leftarrow R2 + 1$, $R5 \leftarrow R5 - 1$
$L14$ IF $0 < R5$, GOTO $L13$
$L15$ IF $0 < R1$, GOTO $L5$
$L16$ $R3 \leftarrow R3 - 1$
$L17$ IF $0 < R3$, GOTO $L16$
$L18$ $R2 \leftarrow R2 - 1$, $R1 \leftarrow R1 + 1$,
$L19$ IF $0 < R2$, GOTO $L18$,

A function is recursive (computable) if and only if it can be computed by a register machine (Minsky [1961] [1967], Melzak [1961], Lambek [1961], Shepherdson

and Sturgis [1963].) The machine M of Example 1 computes the function $f(x) = F_x$, where F_x is the xth Fibonacci number, (Fibonacci number with index x in the enumeration $F_0 = 0$, $F_1 = 1$, $F_{x+2} = F_x + F_{x+1}$). If M is started on line $L1$ with x in register $R1$ and zeros in the other registers, then M eventually halts with $f(x) = F_x$ in register $R1$ and zeros in the other registers. We can understand this as a general definition of computability.

To arithmetize the work of an arbitrary register machine, we will use digits of numbers written to a base Q, where Q is power of 2. Consider the register machine M of Example 1 or more generally any register machine with r registers and l lines in its program. Assume that M computes the total function $y = f(x)$. During the computation the contents of a register at time t can never exceed $x + t$. We will use the fact that after s steps in the computation, the contents of each register Ri is $\leqslant x + s$.

Suppose that after s steps the value $y = f(x)$ is obtained by M. Let $r_{j,t}$ denote the *contents* of register Rj at time t during the course of the computation. If Q is large enough, we will have $r_{j,t} < Q$. So the numbers $r_{j,t}$ may be considered to be the digits of a number written to the base Q. But we will need more room. We will require $2r_{j,t} < Q$, (for the reason see the explanation following (4.19)). Hence we will require that Q satisfy:

$$x + s < Q/2, \tag{4.7}$$

$$l + 1 < Q, \tag{4.8}$$

$$Q \text{ pow } 2. \tag{4.9}$$

For example the number $Q = 2^{x+s+l+1}$ will be large enough. To describe the location of the machine in the code at each time, let $l_{j,t}$ be 1 or 0 according as we execute, or do not execute the instruction(s) at location Lj at time t. When the machine M computes, it will then generate numbers R_j and L_i where

$$R_j = \sum_{t=0}^{s} r_{j,t} Q^t \qquad \left(0 \leqslant r_{j,t} < \frac{Q}{2}\right), \tag{4.10}$$

$$L_i = \sum_{t=0}^{s} l_{i,t} Q^t \qquad (0 \leqslant l_{i,t} \leqslant 1). \tag{4.11}$$

Here the index j runs from 1 to r, where r is the number of *registers*. The index i runs from 1 to l, where l is the number of *lines* in the program.

It will be useful also to have a number I which when written to the base Q has all digits equal to 1. The geometric series will allow us to obtain such a number I.

$$\text{If} \quad 1 + (Q - 1)I = Q^{s+1}, \quad \text{then} \quad I = \sum_{t=0}^{s} Q^t. \tag{4.12}$$

To visualize the arithmetization, consider the machine of Example 1 which has $r = 5$ registers and $l = 19$ lines in its program. When it is started on $x = 2$ as input (i.e. with $x = 2$ in register $R1$), the machine will be seen to halt after $s = 23$ steps, with 1 in register $R1$ and zeros in the other registers. (This is because $F_2 = 1$ is the second Fibonacci number in the standard enumeration $0, 1, 1, 2, 3, 5, 8, 13, \ldots$.) During the computation of F_2, the machine will generate the following numbers R_1, R_2, R_3, R_4, R_5, representing successive contents of the registers, in the sense of (4.10). These numbers, when written in base Q, would

look as follows:

$$R_1 = 11000000000000000011222,$$
$$R_2 = 00111111100000111111100,$$
$$R_3 = 00001122222100001111100,$$
$$R_4 = 00000000001122111000000,$$
$$R_5 = 00000000111111111000000.$$

Similarly, numbers representing the sequence of program locations $L_1, L_2, \ldots, L_{20}$, will be generated during the computation of F_2. In base Q they will look as follows:

$$L_1 = 00000000000000000000001,$$
$$L_2 = 00000000000000000000010,$$
$$L_3 = 00000000000000000000100,$$
$$L_4 = 00000000000000000001000,$$
$$L_5 = 00000000000000000010000,$$
$$L_6 = 00000000000000000100000,$$
$$L_7 = 00000000000000001000000,$$
$$L_8 = 00000000000000010000000,$$
$$L_9 = 00000000000000100000000,$$
$$L_{10} = 00000000000001000000000,$$
$$L_{11} = 00000000000101000000000,$$
$$L_{12} = 00000000010100000000000,$$
$$L_{13} = 00000000100000000000000,$$
$$L_{14} = 00000001000000000000000,$$
$$L_{15} = 00000010000000000000000,$$
$$L_{16} = 00001010000000000000000,$$
$$L_{17} = 00010100000000000000000,$$
$$L_{18} = 00100000000000000000000,$$
$$L_{19} = 01000000000000000000000,$$
$$L_{20} = 10000000000000000000000,$$

Notice that the generated numbers $L_1 L_2, \ldots, L_{20}, R_2, R_3, R_4, R_5$ contain the entire history of the computation of $f(2) = F_2 = 1$.

Now we show how to write down diophantine conditions on arbitrary variables $L_1, L_2, \ldots, L_{20}, R_1, R_2, R_3, R_4, R_5, s, Q, x, y$ sufficient to force them to be these particular values. That is, we give Diophantine conditions on these unknowns which are satisfiable if and only if, on input x, the machine M produces the output value $y = f(x)$. The unknowns in these diophantine relations will be s, Q, $I, R_1, \ldots, R_r, L_1, \ldots, L_{1+1}$. The quantities r and l will be constants, (e.g. $r = 5$ and $l = 19$ in the example). After the conditions have been written down, the methods of §1, §2 and §3 may be used to translate them into diophantine equations in these same unknowns, and more unknowns. The variables x and y will be parameters in the resulting equations.

The first four conditions will be (4.7), (4.8), (4.9) and (4.12). Then to force an arbitrary natural number R_j to have the form (4.10), we will use the masking

relation

$$R_j \preccurlyeq \left(\frac{Q}{2} - 1\right) \cdot I, \qquad (j = 1, \ldots, r). \tag{4.13}$$

Condition (4.13) is implied by condition (4.10), because Q is a power of 2. We also need to force exactly one digit to be 1 in every column of the L matrix. Since $l < Q$ by (4.8), we can use for this purpose the following two conditions:

$$I = \sum_{i=1}^{l+1} L_i, \tag{4.14}$$

$$L_i \preccurlyeq I \qquad (i = 1, \ldots, l+1). \tag{4.15}$$

To start the machine on line $L1$, we stipulate as a starting condition that

$$1 \preccurlyeq L_1. \tag{4.16}$$

By means of GOTO instructions we can suppose there is only one stop command, and that it is located at the end of the program, on line L_{l+1}. Then the condition for stopping the machine after s steps can simply be

$$L_{l+1} = Q^s. \tag{4.17}$$

To simulate each GOTO command (unconditional transfer), Li GOTO Lk, we include a condition

$$QL_i \preccurlyeq L_k. \tag{4.18}$$

This forces the $t+1$st digit of L_k to be 1 whenever the tth digit of L_i is 1. Here $1 \leqslant k \leqslant l+1$. (Line $Ll+1$ is permitted to be the target of a GOTO.) This same idea, (4.18) can be used also for the conditional transfer command (4.4), Li IF $0 < Rj$, GOTO Lk. (Assume $k \neq i+1$ so the command is not trivial.) Here we use two masking conditions:

$$QL_i \preccurlyeq L_k + L_{i+1} \qquad \text{and} \qquad QL_i \preccurlyeq L_k + QI - 2R_j. \tag{4.19}$$

The first masking condition forces transfer toline Li or line $Li+1$. The second decides which one it will be. A diagram explains how this works:

	Q^6						Q^5						Q^4						Q^3						Q^2						Q^1						Q^0					
$I =$	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1.
$Q \cdot I =$	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
$2 \cdot R_j =$	0	0	0	0	0	0	*	*	*	*	*	0	*	*	*	*	*	0	*	*	*	*	*	0	*	*	*	*	*	0	*	*	*	*	*	0	*	*	*	*	*	0.

Here $Q = 2^6$ and $s = 5$ and the idea is that subtraction of $2R_j$ from QI pulls a bit from the Q^{t+1} position of $Q \cdot I$ when $0 < r_{j,t}$. Thus (4.19) implies that $0 < r_{j,t}$ if and only if $l_{k,t+1} = 1$. The reason why we needed $2r_{j,t} < Q$ and why we divided Q by 2 in (4.13) can now be understood. The purpose was to create the zeros between adjacent blocks of digits in $2R_j$. The reason we use $QI - 2R_j$ and not $I - 2R_j$ is that we could not be sure that $I - 2R_j$ would be nonnegative.

The command Li IF $Rj = 0$, GOTO Lk may be treated analogously. This command is just the opposite of (4.4), so it can be simulated with a masking condition like (4.19) but with the second occurrence of L_k replaced by L_{i+1}. This has the effect of causing the machine to go to $Li+1$ instead of Lk, when $Rj > 0$. (Again we are supposing that $k \neq i+1$ so that Li IF $Rj = 0$, GOTO Lk is nontrivial.)

While commands of type $Li\ Rj \leftarrow Rj \pm 1$ are not usually thought of as transfer commands, there is an implied GOTO associated with them; GOTO the next line. So, for each occurrence of such a command (more precisely for each line Li containing occurrences of such commands), we need one corresponding masking condition:

$$QL_i \preccurlyeq L_{i+1}. \tag{4.20}$$

Finally it is necessary to include for each register an equation which ensures that the contents of the register is equal to the tth Q-ary digit of the corresponding number Rj at each time t. We call these *register equations*. For the example program they would be:

$$R_1 + yQ^{s+1} = QR_1 + x + \sum_k QL_k - \sum_i QL_i, \tag{4.21}$$

$$R_j = QR_j + \sum_k QL_k - \sum_i QL_i, \quad (\text{for } j = 2, 3, \ldots, r). \tag{4.22}$$

The k-sums are over all k for which there is an instruction of type $Lk\ Rj \leftarrow Rj + 1$, on line Lk. The i-sums are over all i for which the program has a command of type $Li\ Rj \leftarrow Rj - 1$ on line Li.

The register equation for register $R1$ is different from the register equations for the other registers because $R1$ is the input–output register (assuming functions of 1 variable). At time $t = 0$, register R_1 contains x. At time s it contains y. The other registers contain zeros at these times.

If more registers are designated as input–output registers, for example if $m > 1$, then it is necessary to add a term x to the right side of (4.22) and a term yQ^{s+1} to the left side of (4.22), for the corresponding register equation.

From these equations and conditions, it is not difficult to show, by induction on t_1, that if machine M computes function f, then for all x, y, conditions (4.7), (4.8), (4.9), (4.12)–(4.22) have a solution in the unknowns $s, Q, I, R_1, \ldots, R_r, L_1, \ldots, L_{l+l}$ if and only if $f(x) = y$. Methods of §1, §2, and §3 then permit us to write conditions (4.7), (4.8), (4.9), (4.12)–(4.22) as polynomial equations in the unknowns $s, Q, I, R_1, \ldots, R_r, L_1, \ldots, L_l$, and more unknowns. After we rename the unknowns $x_1, x_2, \ldots, x_n$, transpose all the terms in the equations to one side and sum squares as in (1.5), we obtain a single polynomial equation $P(x, y, x_1, x_2, \ldots, x_n) = 0$ with the property that $(\exists x_1, x_2, \ldots, x_n)$ $[P(x, y, x_1, x_2, \ldots, x_n) = 0]$ if and only if $f(x) = y$.

To complete the proof of Theorem 1, let A be any r.e. set. Suppose $A \neq \varnothing$. Let f be a recursive function whose range is A, so that $y \in A$ iff $(\exists x)\,[f(x) = y]$. Let M be a register machine computing f. When A is a 1-ary relation, $A(a_1)$, we have $a_1 \in A$ iff $(\exists x)\ [f(x) = a_1]$. Thus $a_1 \in A$ if and only if $(\exists x, x_1, x_2, \ldots, x_n)$ $[P(x, a_1, x_1, x_2, \ldots, x_n) = 0]$. When A is an m-ary relation, $A(a_1, a_2, \ldots, a_m)$, then we have $(a_1, a_2, \ldots, a_m) \in A$ iff $(\exists x)\ [f(x) = (a_1, a_2, \ldots, a_m)]$. So P must include m copies of register equation (4.21), with yQ^{s+1} replaced by a_iQ^{s+1} $(i = 1, \ldots, m)$. Then $A(a_1, a_2, \ldots, a_m)$ holds iff $(\exists x, x_1, \ldots, x_n)$ $[P(x, a_1, \ldots, a_m, x_1, \ldots, x_n) = 0]$. Theorem 1 is proved.

§5. Hilbert's 10th problem is unsolvable. We give here a proof that Theorem 1 implies the recursive unsolvability of Hilbert's Tenth Problem. To shorten the proof T. Rado's game [1962] will be used.

THEOREM 2. *There exists no algorithm to solve Hilbert's Tenth Problem.*

Proof. Consider the problem of programming a register machine to return the largest possible number in register $R1$ and stop. Suppose the registers are initially set to 0. Let $R(l)$ be the largest possible number generable in this way, by an l line register machine with registers set initially to 0.

Restricting subtraction commands to the (safe) type (4.1)–(4.2), and disallowing parallelization of commands, $R(l)$ is evidently a well defined function of l. A finite set has a greatest element. There are only a finite number of l line register machines.

Evidently we have $R(1) = 1$ and $R(2) = 2$. Also R is an increasing function of l, $R(l) < R(l+1)$. This can be seen by adding a new line, $R1 \leftarrow R1 + 1$, to the end of an l line program.

Define a binary relation S by putting $S = \{(k,l)\colon k \leqslant R(l)\}$. Then

$$(k,l) \in S \Leftrightarrow (\exists s)[\text{some } l \text{ line machine halts in } s \text{ steps with contents of } R1 \geqslant k]. \tag{5.1}$$

From (5.1) it follows that the set S is r.e. This set S is related to the function R by

$$R(l) = \min k[\neg S(k+1,l)]. \tag{5.2}$$

Apply Theorem 1 to S. The result is a polynomial $P(k,l,x_1,\ldots,x_n)$ with the property that

$$S(k,l) \Leftrightarrow (\exists x_1,x_2,\ldots,x_n)[P(k,l,x_1,x_2,\ldots,x_n) = 0]. \tag{5.3}$$

From (5.2) and (5.3) it follows that

$$R(l) = \min k[\neg(\exists x_1,x_2,\ldots,x_n)[P(k+1,l,x_1,x_2,\ldots,x_n) = 0]]. \tag{5.4}$$

Hence if Hilbert's Tenth Problem were solvable, then R would be computable. But the function R grows very rapidly. In fact it is noncomputable.

LEMMA 5.5. *Let f be any computable function. Then for l sufficiently large, $f(l) < R(l)$.*

Proof. Without loss of generality we can suppose that f is an increasing function. Let N be a register machine which computes the function f. Suppose N has c lines in its program. Let F be the machine obtained by adding the instruction $R1 \leftarrow R1 + 1$ to the end of N. Then F has $c+1$ lines and it computes the function $f(x)+1$.

Let D be a 5 line register machine with the property that when started with x in $R2$ places $2x$ in $R1$ and stops. D could be the machine,

$$\begin{aligned} L1 \quad & \text{IF } R2 = 0, \text{ GOTO } L6, \\ & R2 \leftarrow R2 - 1, \\ & R1 \leftarrow R1 + 1, \\ & R1 \leftarrow R1 + 1, \\ & \text{GOTO } L1, \end{aligned} \tag{5.6}$$

For each x, let M_x be a register machine with x lines and the property that when started with 0 in $R2$ produces x in $R2$. For example M_x could be taken to be a straight line program consisting of x copies of $R2 \leftarrow R2 + 1$.

Now consider the register machine $F(D(M_x))$. By the notation $F(D(M_x))$ is meant M_x followed by D followed by F, in that order. $F(D(M_x))$ is a register machine with $x + 6 + c$ lines in its program. It has the property that for each x, when started with registers initially set to zero, it will produce the number $f(2x) + 1$ in $R1$, then halt. The existence of this machine proves that

$$f(2x) < R(x + 6 + c) \tag{5.7}$$

for each x. When $x \geqslant 6 + c$, we have $x + 6 + c \leqslant 2x$. Hence

$$f(x + 6 + c) \leqslant f(2x), \tag{5.8}$$

because f is increasing. Now put $l = x + 6 + c$. Then for $l \geqslant 12 + 2c$, (5.7) and (5.8) imply $f(l) < R(l)$. Lemma 5.5 is proved. Hilbert's Tenth Problem is unsolvable.

§6. Every computable function is a polynomial. Unsolvability is a negative result, but the solution of Hilbert's Tenth Problem has many positive consequences. One is the existence of polynomials whose set of positive values coincides with any given r.e. set. This result is well known, so we shall prove a stronger form of it which at first glance appears even more surprising.

THEOREM 3. *Let f be any computable function. Then there exists a polynomial Q, with integer coefficients, such that for all nonnegative integers x and y*

$$f(x) = y \leftrightarrow (\exists x_0, x_1, \ldots, x_n)\big[Q(x, x_0, x_1, \ldots, x_n) = y\big]. \tag{6.1}$$

Proof. We apply Theorem 1 to the graph of f (which is an r.e. set), and then use a trick due to H. Putnam [1960]. Starting from the polynomial P which Diophantine defines the graph of f, and using also the fact that f is nonnegative, $(\forall x)\,[f(x) \geqslant 0]$, we get

$$\begin{aligned} f(x) = y &\Leftrightarrow (\exists x_1, \ldots, x_n)\big(P(x, y, x_1, \ldots, x_n) = 0\big), \\ &\Leftrightarrow (\exists x_0, \ldots, x_n)\big(1 - P(x, x_0, x_1, \ldots, x_n)^2 > 0 \text{ and } x_0 = y\big), \\ &\Leftrightarrow (\exists x_0, \ldots, x_n)\Big((x_0 + 1)\big[1 - P(x, x_0, x_1, \ldots, x_n)^2\big] = y + 1\Big). \end{aligned}$$

In the $\Leftarrow$ direction we are using the fact that $1 - P(x, x_0, x_1, \ldots, x_n)^2 > 0$ implies $P(x, x_0, x_1, \ldots, x_n) = 0$ which implies $f(x) = x_0$. So we can put $Q(x, x_0, \ldots, x_n) = (x_0 + 1)[1 - P(x, x_0, x_1, \ldots, x_n)^2] - 1$.

To illustrate Theorem 3, we could apply it to the function $f(x) = F_x$, whose range is the set of Fibonacci numbers. In this case the theorem would give us a polynomial $Q(x, x_0, \ldots, x_n)$ such that for all x, y

$$F_x = y \Leftrightarrow (\exists x_0, \ldots, x_n)\big[Q(x, x_0, \ldots, x_n) = y\big]. \tag{6.3}$$

This would be a Fibonacci number representing polynomial which gives the sequence in order, as a function of x.

The proof of Theorem 1 shows how to construct such a polynomial. If we use this proof, the number n of unknowns will be rather excessive. By another construction, however, it can be shown that we need only nine unknowns. $n = 9$ here and generally in Theorems 1 and 3. This result, the nine unknowns theorem, is a theorem of the second author. A proof can be found in a paper of the first author [1982].

REFERENCES

[1869] H. Anton, Arch. Math. Phys. Vol. 49 (1869), pp. 303–306. (L. E. Dickson, History of the Theory of Numbers, vol. 1, Carnegie Institute of Washington, 1919, 1920, 1923. Chelsea Publishing Co., N.Y., 1971, p. 271.)

[1970] G. V. Chudnovskii, Diophantine predicates (Russian), *Uspekhi Matematicheskikh Nauk*, 25 (1970), 185–186.

[1958] M. D. Davis, Computability and Unsolvability, McGraw-Hill, New York, 1958, xxv + 210 pp.

[1961] M. D. Davis, H. Putnam, and J. Robinson, The decision problem for exponential diophantine equations, *Annals of Math. Series* 2, 74 (1961), 425–436. MR 24, #A3061.

[1971] M. D. Davis, An explicit diophantine definition of the exponential function, *Comm. Pure and Applied Math.*, 24 (1971), 137–145.

[1973] _____, Hilbert's tenth problem is unsolvable, this MONTHLY, 80 (1973), 233–269. MR 47, #6465.

[1976] M. D. Davis, Y. V. Matijasevič & J. Robinson, Hilbert's tenth problem, diophantine equations: positive aspects of a negative solution, Mathematical Developments Arising from the Hilbert problems, Proc. of Symposia in Pure Math., 28, American Math. Soc., Providence, RI, 1976, 323–378. MR55, #5522.

[1980] C. Dimitracopoulos, Matijasevič's Theorem and Fragments of Arithmetic, Dissertation, University of Manchester, 1980.

[1900] D. Hilbert, Mathematische Problem, Vortrag, gehalten auf dem Internationaler Mathematiker Kongress zu Paris 1900, Nachrichten der Akademie der Wissenschaften in Göttingen.II. Mathematische-Physikalische Klasse, 1900, 253–297. English transl: *Bulletin of the American Mathematical Society* 8 (1901–1902), 437–479.

[1974] J. P. Jones, Recursive undecidability—An exposition, MONTHLY 81 (1974), pp. 724–738.

[1976] J. P. Jones, D. Sato, H. Wada and D. Wiens, Diophantine Representation of the Set of Prime Numbers, MONTHLY 83 (1976) 449–464. MR 54, 2615.

[1975] J. P. Jones, Diophantine Representation of the Fibonacci Numbers, *Fibonacci Quarterly*, 13, (1975), 84–88. MR 52, 3035.

[1989] _____, Diophantine Representation of the set of Fibonacci Numbers Over Natural Numbers, Proceedings of the Third International Conference on Fibonacci Numbers and Their Applications, University of Pisa, July 25–29, 1988, Pisa, Italy. Kluwer Academic Publishers.

[1982] _____, Universal diophantine equation, *J. of Symbolic Logic*, 47 (1982), 549–571. MR 84e, #10070.

[1982] J. P. Jones and Y. V. Matijasevič, Exponential diophantine representation of recursively enumerable sets, Proceedings of the Herbrand Symposium, Logic Colloquium '81 (J. Stern, editor), Studies in Logic and the Foundations of Mathematics, 107, North-Holland, Amsterdam, 1982, 159–177. MR 85i, #03138.

[1984] _____, Register machine proof of the theorem on exponential diophantine representation of enumerable sets, *J. of Symbolic Logic*, 49 (1984), 818–829. MR 85i, #03139.

[1971] N. K. Kosovskii, On diophantine representation of the sequence of solutions of Pell's equation (Russian), Zapiski Nauchnyh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V. A. Steklova Akademia Nauk SSSR, 20 (1971), 49–59. English translation: *J. of Soviet Mathematics*, 1 (1973), 28–35.

[1852] E. E. Kummer, Über die Ergänzungssatze zu den Allgemeinen Reziprozitatsgesetzen, *J. für die Reine und Angewandte Mathematik*, 44 (1852), 93–146.

[1770] J. L. Lagrange, Nouv. Mem. Acad. Roy. Sc. de Berlin, 1770, Berlin, 1772, 123–133; *Oeuvres*, 3, 1869, 189–201. See also the edition of Diophantus published by G. Wertheim, 324–330.

[1961] J. Lambek, How to program an infinite abacus, *Canadian Math. Bull.*, 4 (1961), 295–302. MR 24, #A2532.

[1930] D. H. Lehmer, An extended theory of Lucas' functions, *Annals of Math* 31 (1930), 419–448.

[1877] E. Lucas, Sur les congruences des nombres eulériens et des coefficients différentiels des fonctions trigonométriques suivant un module premier, *Bulletin de la Société Mathématique de France*, 6 (1877–78), 49–54.

[1878] _____, Theorie des fonctions numériques simplement périodiques, *American Journal of Mathematics*, 1 (1878), 184–240, 289–321. English translation: The Fibonacci Association, Santa Clara University, CA 95053 (1969).

[1970] Y. V. Matijasevič, Enumerable sets are diophantine, *Doklady Akademii Nauk SSSR*, 191 (1970),

279–282. English translation with addendum, *Soviet Mathematics*: Doklady, 11 (1970), 354–357. MR 41, #3390.

[1971] _____, Diophantine representation of enumerable predicates, *Izvestija Akademii Nauk SSSR Serija Mathematiceskaja*, 35 (1971), 3–30; English Transl. *Mathematics of the USSR–Izvestija*, 5 (1971), 1–28. MR 43, #54.

[1972] _____, Diophantine sets, *Uspekhi Matematičeskih Nauk*, 27 (1972), no. 5 (167), 185–222. English translation, *Russian Mathematical Surveys*, 27 (1972), No. 5, 124–164. MR 56, #109.

[1975] Y. V. Matijasevič and J. Robinson, Reduction of an arbitrary diophantine equation to one in 13 unknowns, *Acta Arithmetica*, 27 (1975), 521–553. MR 52, #8033.

[1984] Y. V. Matijasevič, On investigations of some algorithmic problems in algebra and number theory, in Algebra, Math. Logic, Theory of Numbers and Topology, 50 years of M.I.A.N., Trudy Mathematicheskogo Instituta Steklova, Nauka, 168 (1984), pp. 218–235. English translation: American Mathematical Society, Proceedings of the Steklov Institute (1986), No. 3, pp. 227–252.

[1961] Z. A. Melzak, An informal arithmetical approach to computability and computation, *Canadian Math. Bull.*, 4 (1961), 279–294. MR 27, #1364.

[1961] M. Minsky, Recursive unsolvability of Post's problem of tag and other topics in the theory of Turing machines, *Annals of Mathematics*, Ser. 2, 74 (1961), 437–455. MR 25, #3825.

[1967] _____, Computation: Finite and infinite machines, Prentice-Hall, Englewood Cliffs, New Jersey, 1967. MR 50, #9050.

[1971] R. Parikh, Existence and feasibility in arithmetic, *J. Symbolic Logic*, 36 (1971), 494–508.

[1982] J. B. Paris and C. Dimitracopoulos, Truth definitions for Δ_0 formulae, Logic and Arithmetc, Monographie No. 30, L'Enseignement Mathematique, Engeler, Lauchli & Strassen eds., University of Geneva, 317–330.

[1960] H. Putnam, An unsolvable problem in number theory, *J. of Symbolic Logic* 25 (1960), 220–232.

[1962] T. Rado, On noncomputable functions, *Bell Sys. Tech. Jour.*, 41 (1962) 877–884.

[1952] J. Robinson, Existential definability in arithmetic, *Trans. Amer. Math. Soc.*, 72 (1952), 437–449. MR 14, #4.

[1969] _____, Diophantine decision problems, Studies in number theory (W. J. LeVeque, editor), MAA Studies in Mathematics, 6, Mathematical Association of America, Buffalo, New York (distributed by Prentice–Hall, Englewood Cliffs, New Jersey), 1969, 76–116. MR 39, #5364.

[1967] H. Rogers Jr., Theory of Recursive Functions and Effective Computability, McGraw-Hill, New York, 1967, ix + 482 pp.

[1963] J. C. Shepherdson and H. E. Sturgis, Computability of recursive functions, *J. of the Assoc. for Computing Machinery*, 10 (1963), 217–255. MR 27, #1359.

Linear Algebra and Quantum Chemistry

JEROME A. GOLDSTEIN AND MEL LEVY, *Tulane University*

JERRY GOLDSTEIN received all his degrees from Carnegie-Mellon University, obtaining his Ph.D. in 1967. After a year at The Institute for Advanced Study he went to Tulane. Among his visiting positions are Brasilia, Federal University of Rio de Janeiro, London, Padova, Graz, and Tübingen. His research interests are operator theory, partial differential equations, and quantum mechanics. Among his publications is the monograph *Semigroups of Linear Operators and Applications*, Oxford, 1985.



MEL LEVY received his Ph.D. in chemical physics, with a specialty in quantum chemistry, at Indiana University in 1971. He has been on the faculty of Tulane University since 1976. His research interests have focused upon the formal foundations of both quantum chemistry and density-functional theory.



1. Introduction. Our goal is to show how some problems in quantum chemistry arise naturally as minimization problems involving $N \times N$ matrices in the Hilbert-Schmidt norm. These minimization problems have the added feature that the solutions remain solutions when the Hilbert-Schmidt norm is replaced by any other of the "usual" norms. One outgrowth of our discussions will be a scientifically motivated procedure for converting a basis into an orthogonal one; the Löwdin procedure will arise "naturally" and will have a nice property not shared by the better-known Gram-Schmidt method.

In Section 2 we review some aspects of the theory of linear operators on a finite dimensional Hilbert space. This is the "linear algebra" part of this paper. Sections 3 through 6 contain the main results.

2. Finite dimensional operator theory. Let $(\mathscr{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ be a finite dimensional complex Hilbert space. Let N be its dimension. We may suppose for concreteness that $\mathscr{H} = \mathbf{C}^N$. Let $\mathscr{B}(\mathscr{H})$ be all the linear operators on $\mathscr{H}$; that is, $\mathscr{B}(\mathscr{H}) = M_N(\mathbf{C})$ consists of all $N \times N$ complex matrices. For $A \in \mathscr{B}(\mathscr{H})$ let $|A| = (A^*A)^{1/2}$. By the spectral theorem,

$$|A| = \sum_{n=1}^{N} \lambda_n e_n \otimes e_n = \sum_{n=1}^{N} \lambda_n \langle \cdot, e_n \rangle e_n, \tag{2.1}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant 0$ are the eigenvalues and $e_1, \ldots, e_N$ are the corresponding normalized eigenvectors of $|A|$. Here $e_n \otimes e_n$ is the orthogonal projection onto the span of e_n. Because $\|Af\|^2 = \langle Af, Af \rangle = \langle A^*Af, f \rangle = \||A|f\|^2$, we can write

$$A = U|A|, \tag{2.2}$$

where U is a unitary operator from the range of $|A|$ to the range of A. The decomposition (2.2) is called the *polar decomposition* of A (cf., e.g. [7]). It reduces to $z = e^{i\theta}|z|$ in $\mathbf{C}^1$. When A is injective, U is unitary on $\mathscr{H}$. But U can always be

viewed as being a unitary operator on $\mathscr{H}$ by extending it in any way as a unitary operator from $\mathrm{Ran}(|A|)^{\perp} = \mathrm{Ker}(|A|)$ to $\mathrm{Ran}(A)^{\perp} = \mathrm{Ker}(A^*)$. However, U is uniquely determined as a unitary operator on $\mathscr{H}$ only when A is injective.

Let J_A be the largest index n such that $\lambda_n > 0$ for $n \leqslant J_A$. (Thus $J_O = 0$; $J_A = N$ if and only if A is injective; and $2 \leqslant J_A \leqslant N - 1$ in all other cases.) Using (2.1) and (2.2) we can write A uniquely in the form

$$A = \sum_{n=1}^{J_A} s_n(A)(f_n \otimes e_n) = \sum_{n=1}^{J_A} s_n(A)\langle \cdot\,, e_n\rangle f_n, \tag{2.3}$$

where $s_n(A) = \lambda_n$ is the *nth singular number* of A and $f_n = Ue_n$. Replacing J_A by N would make no difference since $s_n(A) = 0$ for $J_n < n \leqslant N$. The vector $s(A) = (s_1(A), \ldots, s_N(A))$ is called the *singular sequence* of A.

Let E be $\mathbf{C}^N$ made into a Banach space under a norm ν which satisfies

$$\nu\big((x_1, \ldots, x_N)\big) = \nu(\varepsilon_1 x_{\pi 1}, \ldots, \varepsilon_N x_{\pi N})$$

for every $(x_1, \ldots, x_N) \in \mathbf{C}^N$, every permutation π of $\{1, \ldots, N\}$, and every choice of $\varepsilon_j \in \mathbf{C}$ such that $|\varepsilon_j| = 1$ for $1 \leqslant j \leqslant N$. Now make $\mathscr{B}(\mathbf{C}^N)$ into a Banach space, denoted by $\mathscr{C}_E$, under the norm

$$\|A\|_{\mathscr{C}_E} = \nu(s(A)).$$

These spaces were introduced by Schatten and von Neumann (cf. [**12**]). When $E = l_N^p$, so that $\nu((x_1, \ldots, x_N)) = \{\sum_{n=1}^N |x_n|^p\}^{1/p}$ or $\max_n |x_n|$, according as $1 \leqslant p < \infty$ or $p = \infty$, the corresponding space $\mathscr{C}_E$ is usually denoted by $\mathscr{C}_p$. Thus for $1 \leqslant p < \infty$,

$$\|A\|_{\mathscr{C}_p} = \left\{ \sum_{n=1}^{J_A} s_n(A)^p \right\}^{1/p}$$

and $\|A\|_{\mathscr{C}_\infty} = s_1(A) = \|A\| = \sup\{\|Af\|\colon \|f\| \leqslant 1\}$, which is the usual norm on $\mathscr{B}(\mathscr{H})$.

The norms in the spaces $\mathscr{C}_E$ are all *unitarily invariant* in the sense that

$$\|UAV\|_{\mathscr{C}_E} = \|A\|_{\mathscr{C}_E}$$

whenever U and V are unitary. The space $\mathscr{C}_2$ is a Hilbert space under the inner product

$$\langle A, B\rangle_2 = \mathrm{tr}(B^*A),$$

where "tr" stands for trace. The norm in $\mathscr{C}_2$ is called the *Hilbert-Schmidt norm*. For $A = (a_{ij})$ (with respect to a fixed orthonormal basis),

$$\|A\|_{\mathscr{C}_2}^2 = \langle A, A\rangle_2 = \sum_{i,j=1}^{N} |a_{ij}|^2.$$

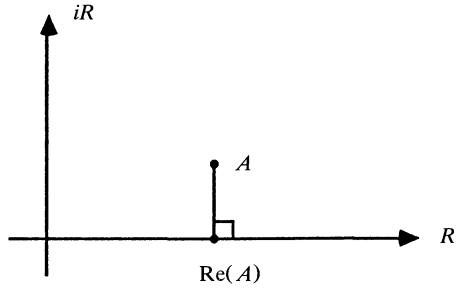
Finally we shall have occasion to refer to *Mirsky's inequality* [**11**], which states that for $A, B \in \mathscr{B}(\mathscr{H})$,

$$\sum_{n=1}^{m} |s_n(A) - s_n(B)| \leqslant \sum_{n=1}^{m} |s_n(A - B)|$$

whenever $1 \leqslant m \leqslant N$. (For infinite dimensional generalizations involving compact operators, see Markus [**10**].)

3. A computational problem from quantum theory. Suppose that a chemist gives us an operator A in $\mathscr{B}(\mathscr{H})$ (with $\mathscr{H} = \mathbf{C}^N$ as before). We are told that A is a finite dimensional approximation of a Hamiltonian and we are asked to find its ground state (or lowest) eigenvalue and eigenvector. We note that A is not self-adjoint. The chemist says "sorry," this is a result of some small errors due to the computer, so we are instructed to replace A by the self-adjoint operator of which it is a small perturbation. But which operator is that? Seemingly we are to find the self-adjoint operator nearest to A in some reasonable sense.

Picture our arbitrary operator A in $\mathscr{B}(\mathscr{H})$ as being represented by a point in the complex plane, and think of the real line as representing the self-adjoint operators on $\mathscr{H}$.



The picture suggests that the nearest self-adjoint operator to A is $\mathrm{Re}(A) = (1/2)(A + A^*)$. This is in fact true if we measure the distance between operators by any of the unitarily invariant norms introduced in Section 2. The proof is as follows. If $S = S^*$ in $\mathscr{B}(\mathscr{H})$, then

$$\|A - \mathrm{Re}(A)\|_{\mathscr{C}_E} = \left\|\tfrac{1}{2}(A - S) + \tfrac{1}{2}(S - A^*)\right\|_{\mathscr{C}_E}$$
$$\leqslant \tfrac{1}{2}\|A - S\|_{\mathscr{C}_E} + \tfrac{1}{2}\|S - A^*\|_{\mathscr{C}_E} = \|A - S\|_{\mathscr{C}_E}$$

since $\|B^*\|_{\mathscr{C}_E} = \|B\|_{\mathscr{C}_E}$. QED.

When is $\mathrm{Re}(A)$ the *unique* solution? Answer: Whenever the norm in $\mathscr{C}_E$ is strictly convex.[†] Moreover, the above diagram is a truly correct picture in the Hilbert-Schmidt case. Think of $\mathscr{C}_2$ as a *real* Hilbert space of dimension $2N^2$. The self-adjoint operators on $\mathscr{H}$ form a real subspace $\mathscr{S}$ of $\mathscr{C}_2$ of dimension N^2. The orthogonal projection $P_{\mathscr{S}}$ of $\mathscr{C}_2$ onto $\mathscr{S}$ is given by $P_{\mathscr{S}} = \mathrm{Re}$, i.e.,

$$P_{\mathscr{S}}(A) = \tfrac{1}{2}(A + A^*).$$

Thus $P_{\mathscr{S}}$ is the unique nearest point in $\mathscr{S}$ to A since the orthogonal projection is also the nearest point map. Thus the above picture is justified.

$\mathrm{Re}(A)$ is the *unique* solution in $\mathscr{C}_p$ for $1 < p < \infty$, but uniqueness fails in $\mathscr{C}_1$ and in $\mathscr{C}_\infty$ ($= \mathscr{B}(\mathscr{H})$ with its "usual" norm).

The above results generalize and simplify the discussion in [**5**].

[†]A norm $|\cdot|$ is strictly convex whenever $|x + y| = |x| + |y|$ implies that either $x = ty$ or $y = tx$ for some $t \geqslant 0$.

4. The Löwdin problem. Theoretical chemistry considerations associated with basis set optimization and with the understanding of properties in terms of electron pair (chemical bond) transferability lead to the following mathematical problem.[†]

We are given a basis $\{f_1, \ldots, f_N\}$ of unit vectors for $\mathscr{H}$. (In quantum mechanics all states are represented by unit vectors.) These vectors contain much information (namely, the information from the known subsystems). In order to do theoretical and computational studies it is necessary to orthogonalize the $\{f_n\}$. This is done by an *orthogonalization*, that is, an operator $B \in \mathscr{B}(\mathscr{H})$ such that $Bf_n = e_n$ satisfies $\langle e_n, e_m \rangle = \delta_{mn} = 1$ or 0 according as $m = n$ or $m \neq n$, for $1 \leqslant m, n \leqslant N$. But one wishes to change the $\{f_n\}$ as little as possible because changing the $\{f_n\}$ results in a loss of information. How can B be chosen in an optimal way?

A natural approach to this problem is to choose B to minimize the least squares deviation

$$\sum_{n=1}^{N} \|f_n - e_n\|^2.$$

Writing $B = UL$ (with $L = |B|$) as the polar decomposition for the invertible operator B we have

$$\begin{aligned}\sum_{n=1}^{N} \|f_n - e_n\|^2 &= \sum_{n=1}^{N} \|B^{-1}e_n - e_n\|^2 \\ &= \sum_{n=1}^{N} \|(L^{-1} - U)U^{-1}e_n\|^2 = \|L^{-1} - U\|_{\mathscr{C}_2}^2,\end{aligned}$$

since

$$\|A\|_{\mathscr{C}_2}^2 = \operatorname{tr}(A^*A) = \sum_{n=1}^{N} \langle A^*Ag_n, g_n \rangle = \sum_{n=1}^{N} \|Ag_n\|^2$$

holds for *any* orthonormal basis $\{g_1, \ldots, g_N\}$.

Replace B by another orthogonalization $\tilde{B}$, where $\tilde{B}f_n = \tilde{e}_n$ for all n and $\langle \tilde{e}_n, \tilde{e}_m \rangle = \delta_{mn}$ for all m, n. Since

$$f_n \xrightarrow{UL} e_n \xrightarrow{V} \tilde{e}_n$$

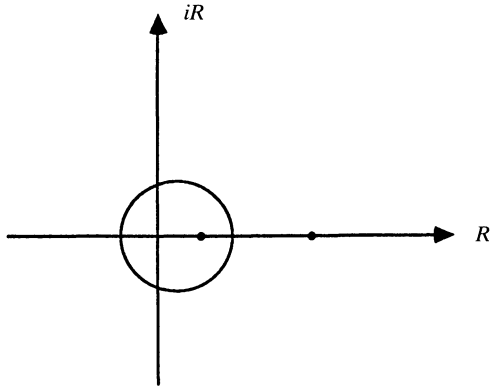
and the linear map V from e_n to $\tilde{e}_n$ for all n is a unitary operator, the invertible $\tilde{B}$ is given by $V(UL) = (VU)L$, which is its unique polar decomposition. Thus $L = |B| (= |\tilde{B}|)$ is independent of the orthogonalization B, and varying $B = UL$ is equivalent to varying U over $\mathscr{U}(N)$, the group of all unitary operators on $\mathbf{C}^N$. Thus

$$\min_{B} \sum_{n=1}^{N} \|f_n - e_n\|^2 = \min_{U \in \mathscr{U}(N)} \|L^{-1} - U\|_{\mathscr{C}_2}^2. \tag{4.1}$$

We now "solve" this minimization problem *pictorially* (as in Section 4). In the figure below, the unit circle represents the unitary operators in $\mathscr{U}(N)$ and the positive real axis represents the positive self-adjoint operators. Thus L^{-1} is represented by one of the fat dots. Geometrically the point on the unit circle

[†]The point is that real issues in quantum chemistry motivate our mathematical problem. For some of the motivation, see [3], [5], [8].

closest to either dot is the point $(1, 0)$. Thus we conjecture that the minimum (4.1) is achieved for $B = L$ or, equivalently, for $U = I$. This is in fact correct.



There are several ways to prove this. Here is a calculus proof taken from [1], [2]. Let $A = A^*$ be a fixed positive self-adjoint operator on $\mathscr{H} = \mathbf{C}^N$. Let

$$F(U) = \|A - U\|_{\mathscr{C}_2}^2;$$

F: $\mathscr{U}(N) \to [0, \infty)$ is a smooth map on a compact manifold. If it has a local minimum at $U_0 \in \mathscr{U}(N)$, then $F'(U_0) = 0$. Let P be any rank one self-adjoint projection on $\mathscr{H}$ and let $f(t) = F(e^{itP}U_0)$. Then since the smooth function f: $\mathbf{R} \to \mathbf{R}$ has a local minimum at $t = 0$, we must have $f'(0) = 0$, $f''(0) \geqslant 0$. Using

$$f'(t) = \frac{d}{dt}\operatorname{tr}\left[\left(A - e^{itP}U_0\right)^*\left(A - e^{-itP}U_0\right)\right],$$

and differentiating inside the trace symbol, a short calculation (cf. [1]) yields the following restrictions on the spectrum of U_0:

$$f'(0) = 0 \text{ for all } P \text{ implies } \sigma(U_0) \subset \mathbf{R};$$

$$f''(0) \geqslant 0 \text{ for all } p \text{ implies } \operatorname{Re}\sigma(U_0) \subset [0, \infty).$$

Thus, since U_0 is unitary, $\sigma(U_0) = \{1\}$ and $U_0 = I$.

Here is a recipe for finding L (or L^{-1}). Recall that $\{f_1, \ldots, f_N\}$ is given. The Gram matrix is the $N \times N$ matrix $G = (g_{ij})$ defined by $g_{ij} = \langle f_i, f_j \rangle$. This is a positive definite matrix and $L = G^{-1/2}$ (or $L^{-1} = G^{1/2}$).

The calculus method used above works for other norms as well. Here is a sample result. Let $1 \leqslant p \leqslant \infty$ and let

$$F_p(U) = \|A - U\|_p,$$

where A is a given positive self-adjoint operator on $\mathbf{C}^N$ and U varies over $\mathscr{U}(N)$. Then

$$\min_U F_p(U) = F_p(I)$$

for $1 \leqslant p \leqslant \infty$. Moreover, when $1 < p < \infty$, $U = I$ is the only local minimum of F_p (but this is not the case when $p \in \{1, \infty\}$ and $N \geqslant 2$). See [2] for more details and for infinite dimensional versions. See also Section 6 for additional results.

The minimization problem solved in this section goes back to the Swedish quantum chemist Per-Olof Löwdin [**9**]. That's why we used the letter L. The *Löwdin orthogonalization* (i.e., $B = L$) is a *symmetric* orthogonalization in two senses. First L is symmetric, i.e., self-adjoint. Next, if the given basis $\{f_1, \ldots, f_N\}$ is permuted to become $\{f_{\pi 1}, \ldots, f_{\pi N}\}$, then the same $B = L$ results, and the new "nearest" orthonormal basis we obtain is $\{e_{\pi 1}, \ldots, e_{\pi N}\}$ where $\{e_1, \ldots, e_N\}$ is the original Löwdin orthonormal basis (i.e., $e_n = Lf_n$ for all n). So the Löwdin orthogonalization procedure is symmetric with respect to permutations. This is a clear indication of how the Löwdin procedure differs from the Gram-Schmidt orthogonalization procedure.

As a pedagogical aside we mention that we advocate the teaching of the Löwdin orthogonalization method in linear algebra courses since (among other reasons) it is easy to motivate by physical considerations.

5. A more general approximation problem. Let $\{f_1, \ldots, f_N\}$ be a given basis of unit vectors for $\mathscr{H} = \mathbf{C}^N$. It is desired to replace $\{f_n\}$ by an orthonormal basis $\{e_1, \ldots, e_N\}$ such that

(i) $\{e_1, \ldots, e_N\}$ is close to $\{f_1, \ldots, f_N\}$,

(ii) $\{e_1, \ldots, e_N\}$ is close to a given orthonormal basis $\{g_1, \ldots, g_N\}$.

If constraint (ii) is missing, this reduces to the Löwdin problem stated and solved in the previous section. Condition (ii) may be present to test a hypothesis that some competing chemical theory is right (or wrong). For more background and a fuller discussion see [**5**].

Let $0 \leqslant b \leqslant 1$. Let $Bf_n = e_n$ for $n = 1, \ldots, N$, where $\{e_n\}$ is an orthonormal basis. Define

$$F_b(B) = b \sum_{n=1}^{N} \|f_n - e_n\|^2 + (1 - b) \sum_{n=1}^{N} \|g_n - e_n\|^2.$$

F_1 is the Löwdin functional previously considered. Decreasing b makes constraint (ii) more important relative to (i).

We want to minimize $F_b(\mathscr{B})$ as $\mathscr{B}$ varies (with b fixed). For $b = 0, 1$ we know the answer, although the problem is of no physical (or mathematical) interest when $b = 0$. As before write $B = UL$ for the unique polar decomposition of the orthogonalization B. The particular orthogonalization sending f_n to g_n for each n we denote by WL. Then

$$\begin{aligned} F_b(B) &= b \sum_{n=1}^{N} \|L^{-1}U^{-1}e_n - e_n\|^2 + (1 + b) \sum_{n=1}^{N} \|WLf_n - ULf_n\|^2 \\ &= b \sum_{n=1}^{N} \|(L^{-1} - U)U^{-1}e_n\|^2 + (1 - b) \sum_{n=1}^{N} \|(W - U)Lf_n\|^2 \\ &= b\|L^{-1} - U\|_{\mathscr{C}_2}^2 + (1 - b)\|W - U\|_{\mathscr{C}_2}^2. \end{aligned} \tag{5.1}$$

Minimizing $F_b(B)$ as B varies is the same as minimizing the right-hand side of (5.1) as U varies over $\mathscr{U}(N)$. In this case the positive self-adjoint operator L^{-1} and the unitary operator W are given and fixed.

A more general problem can be formulated as follows. Given $A_1, \ldots, A_M$ in $\mathscr{B}(\mathscr{H})$ and $b_1, \ldots, b_M$ in $\mathbf{R}$ such that $b_n \geqslant 0$ and $\Sigma_{n=1}^{M} b_n = 1$, minimize

$$G_b(U) = \sum_{n-1}^{M} b_n \|A_n - U\|_{\mathscr{C}_2}^2$$

as U varies over $\mathscr{U}(U)$. The calculus approach of the previous section can be used (cf., [5]). Here is a simple direct Hilbert space argument.

By the law of cosines,

$$G_b(U) = \sum_{n=1}^{M} b_n \left\{ \|A_n\|_{\mathscr{C}_2}^2 + \|U\|_{\mathscr{C}_2}^2 - 2 \operatorname{Re} \operatorname{tr}(U^* A_n) \right\}$$

$$= N + \sum_{n=1}^{M} b_n \|A_n\|_{\mathscr{C}_2}^2 - 2 \sum_{n=1}^{M} b_n \operatorname{Re} \operatorname{tr}(U^* A).$$

Thus minimizing $G_b(U)$ is equivalent to maximizing

$$H_b(U) = \operatorname{Re} \operatorname{tr}(U^* C),$$

where $C = \Sigma_{n=1}^{M} b_n A_n$ (since $G_b(U) + 2H_b(U)$ is a constant, independent of U). Let $C = V|C|$ be the polar decomposition of C. (Note that V need not be uniquely determined as a unitary operator since C need not be injective.) The maximum value of $H_b(U)$ is clearly $\|C\|_{\mathscr{C}_1}$; it is attained whenever $U = V$ on the range of C. Thus the minimization problem for G_b is solved, uniquely so when C is injective.

In the context of $F_b(B)$ as in (5.1), C becomes

$$C = bL^{-1} + (1 + b)W.$$

If $LW = WL$, then C is a normal operator with spectral decomposition

$$C = \sum_{n-1}^{N} \eta_n h_n \otimes h_n$$

for some orthonormal basis $\{h_1, \ldots, h_n\}$ and complex constants $\{\eta_1, \ldots, \eta_n\}$. The minimizing U for $F_b(B)$ is given by

$$Uh_n = \zeta_n h_n$$

for $n \in K = \{n \in \{1, \ldots, N\} : \eta_n \neq 0\}$, where $\zeta_n = \eta_n / |\eta_n|$, while $\{Uh_n : n \in \{1, \ldots, N\} \setminus K\}$ can be an arbitrary orthonormal basis for $\{h_n : n \in K\}^{\perp}$.

6. The Arazy approach. In the fall of 1984, Jonathan Arazy (of Haifa, Israel) visited Tulane University and made a number of interesting observations [4] on Löwdin orthogonalization and related topics. Part of Section 5 is due to him; in particular he noted that the minimum *value* of $F_b(B)$ can be given explicitly in terms of $\mathscr{C}_1$ and $\mathscr{C}_2$ norms.

More interestingly, Arazy noted that Mirsky's inequality (see Section 2) implies that $U = I$ gives the solution of Löwdin problem in *every* unitarily invariant norm $\|\cdot\|_{\mathscr{C}_E}$.

Let ν, E be as in Section 2. Let $A \in \mathscr{B}(\mathbf{C}^N)$ and let $A = U|A|$ be its polar decomposition. For any $V \in \mathscr{U}(N)$,

$$\begin{aligned}
\|A - V\|_{\mathscr{C}_E} &= \nu(s(A - V)) \quad \text{by definition} \\
&\geqslant \nu(s(A) - s(V)) \quad \text{by Mirsky's inequality} \\
&= \nu\big((s_1(A) - 1, \ldots, s_N(A) - 1)\big) \quad \text{since } s(V) = (1, \ldots, 1) \\
&= \left\| \begin{pmatrix} s_1(A) & & \bigcirc \\ & \ddots & \\ \bigcirc & & s_N(A) \end{pmatrix} - \begin{pmatrix} 1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & 1 \end{pmatrix} \right\|_{\mathscr{C}_E} \\
&= \|A - U_0\|_{\mathscr{C}_E}; \qquad (6.1)
\end{aligned}$$

here, as before, $\{e_n\}$ is the orthonormal basis of eigenvectors of $|A|, \{f_n\} = \{Ue_n\}$, and the matrices in (6.1) are taken with respect to the row basis $\{f_n\}$ and the column basis $\{e_n\}$. This way of writing A is nothing but a restatement of (2.3). Thus

$$\|A - V\|_{\mathscr{C}_E} \geqslant \|A - U_0\|_{\mathscr{C}_E}$$

holds for all $V \in \mathscr{U}(N)$ and for all $U_0 \in \mathscr{U}(N)$ which exchange the $\{e_n\}$ with the $\{f_n\}$, in order, in the polar decomposition of A. When A is injective, U_0 is thus uniquely determined; in fact the polar decomposition of A is $A = U_0|A|$. This shows that the Löwdin solution is valid for *all* the norms $\mathscr{C}_E$ on $\mathscr{B}(\mathscr{H})$.

When A is injective, is U_0 the unique solution of the minimization problem in the $\mathscr{C}_E$ norm? This depends upon an analysis of when equality holds in Mirsky's inequality. In particular, this depends on ν.

This approach via Mirsky's inequality can be adapted to the problem of Section 5 as well. For instance, suppose we want to minimize

$$\sum_{n=1}^{M} b_n F_\nu(\|A_n - U\|_{\mathscr{C}_E})$$

as U varies over $\mathscr{U}(N)$. Here $b_n \geqslant 0, \sum_{n=1}^{M} b_n = 1$, and F_ν is some suitable functional related to (E, ν). For instance, when $\mathscr{C}_E = \mathscr{C}_p$ for $1 \leqslant p < \infty$, take $F_\nu(s) = |s|^p$. When $p = 2$, this reduces to the problem of Section 6. For $1 < p < \infty$, we get a *unique* solution to the analogous problem (when $C = \sum_{n=1}^{M} b_n A_n$ is injective), and the $\mathscr{C}_p$ solution coincides with the $\mathscr{C}_2$ solution.

Acknowledgements. This work was completed and prepared during an exhilarating 1988 visit of Goldstein to the Mathematics Institute of the University of Tübingen. He is most grateful for the extraordinary hospitality and the congenial mathematical atmosphere provided by Rainer Nagel and his colleagues. The research was partially supported by the NSF and the DFG. Finally we thank Jonathan Arazy for generously allowing the inclusion of his results in this paper.

REFERENCES

1. J. G. Aiken, J. A. Erdos, and J. A. Goldstein, On Löwdin orthogonalization, *Internat. J. Quantum Chem.*, 18 (1980) 1101–1108.
2. ———, Unitary approximation of positive operators, *Ill. J. Math.*, 24 (1980) 61–72.
3. T. L. Allen, Bond energies and the interactions between next-nearest neighbors. I. Saturated hydrocarbons, diamond, sulfanes, S_8, and organic sulfur components. *J. Chem. Phys.*, 31 (1959) 1039–1049.
4. J. Arazy, personal communication, 1984.

5. M. J. Dupré, J. A. Goldstein, and M. Levy, The nearest self-adjoint operator, *J. Chem. Phys.*, 72 (1980) 780–781.
6. J. A. Goldstein and M. Levy, Hilbert-Schmidt approximation problems arising in quantum chemistry, *Adv. Appl. Math.*, 5 (1984) 216–225.
7. P. R. Halmos, *A. Hilbert Space Problem Book*, Van Nostrand, Princeton, N.J., 1967.
8. M. Levy, W. J. Stevens, H. Schull, and S. Hagstrom, Transferability of electron pairs between H_2O and H_2O_2, *J. Chem. Phys.*, 61 (1924) 1844–1856.
9. P.-O. Löwdin, On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals, *J. Chem. Phys.*, 18 (1950) 365–374.
10. A. S. Markus, Eigenvalues and singular values of the sum and product of linear operators, *Russian Math. Surveys*, 19 (4) (1964) 91–120.
11. L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Quart. J. Math. Oxford*, 11 (1960) 50–59.
12. R. Schatten, *Norm Ideals of Completely Continuous Operators*, Springer, Berlin, 1960.

INTEGER SEQUENCES

For a second edition of "A Handbook of Integer Sequences," new sequences, extensions and corrections are requested. Please supply exact references if possible. N. J. A. Sloane, Room 2C-376, AT & T Bell Laboratories, Murrary Hill, NJ 07974 USA; electronic mail address: njas@research.att.com

The Fifty-First William Lowell Putnam Mathematical Competition

Leonard F. Klosinski, *Santa Clara University*
Gerald L. Alexanderson, *Santa Clara University*
Loren C. Larson, *St. Olaf College*

The following results of the fifty-first William Lowell Putnam Mathematical Competition, held on December 1, 1990, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, $5,000, was awarded to the Department of Mathematics of Harvard University. The members of the winning team were: Jordan S. Ellenberg, Raymond M. Sidney, and Eric K. Wepsic; each was awarded a prize of $250.

The second prize, $2,500, was awarded to the Department of Mathematics of Duke University. The members of the winning team were Jeanne A. Nielsen, Will A. Schneeberger, and Jeffrey M. Vanderkam; each was awarded a prize of $200.

The third prize, $1,500, was awarded to the Department of Mathematics of the University of Waterloo. The members of the winning team were Dorian Birsan, Daniel R. L. Brown, and Colin M. Springer; each was awarded a prize of $150.

The fourth prize, $1,000, was awarded to the Department of Mathematics of Yale University. The members of the winning team were Thomas Zuwei Feng, Andrew H. Kresch, and Zhaoliang Zhu; each was awarded a prize of $100.

The fifth prize, $500, was awarded to the Department of Mathematics of Washington University. The members of the winning team were William Chen, Adam M. Costello, and Jordan A. Samuels; each was awarded a prize of $50.

The five highest ranking individual contestants, in alphabetical order, were Jordan S. Ellenberg, Harvard University; Jordan Lampe, University of California, Berkeley; Raymond M. Sidney, Harvard University; Ravi D. Vakil, University of Toronto; and Eric K. Wepsic, Harvard University. Each of these was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of $500 by the Putnam Prize Fund.

The next seven highest ranking individuals, in alphabetical order, were Andrew H. Kresch, Yale University; Samuel A. Kutin, Harvard University; Royce Y. Peng, Harvard University; Eric M. Rains, Case Western Reserve University; Jeffrey M. Vanderkam, Duke University; Samuel K. Vandervelde, Swarthmore College; and Michael E. Zieve, Harvard University. Each was awarded a prize of $250.

The following teams, named in alphabetical order, received honorable mention: California Institute of Technology, with team members Tien-Yee Chiu, Russell A. Manning, and Robert G. Southworth; the University of California, Berkeley, with team members Brian J. Birgen, Jordan Lampe, and Thomas S. Lumley; Massachusetts Institute of Technology, with team members Christos Athanasiadis, Andrew Chou, and David B. Wilson; Stanford University, with team members, Daniel P. Cory, Gregory G. Martin, and Andras Vasy; and Swarthmore College, with team members Olaf A. Holt, Robert E. Marx, and Samuel K. Vandervelde.

Honorable mention was achieved by the following thirty-five individuals named in alphabetical order: Eric M. Boesch, University of Maryland, College Park;

Hubert L. Bray, Rice University; Daniel R. L. Brown, University of Waterloo; Michael J. Callahan, Harvard University; David B. Carlton, Harvard University; Helmut R. Celina, Davidson College; Nickolai I. Chavdarov, Brandeis University; William Chen, Washington University, St. Louis; Mark T. Chrisman, University of California, Davis; Bryan F. Clair, University of California, Berkeley; Brian D. Conrad, Harvard University; Daniel B. Finn, University of Rochester; Joshua B. Fischman, Princeton University; Mikhail Grinberg, Massachusetts Institute of Technology; Alex Gurevich, University of Maryland, College Park; Richard S. Kiss, Simon Fraser University; John C. Loftin, Stanford University; Gregory G. Martin, Stanford University; David K. McKinnon, Harvard University; Jeanne A. Nielson, Duke University; David M. Patrick, Carnegie Mellon University; Alexander R. Pruss, University of Western Ontario; Jordan A. Samuels, Washington University, St. Louis; Will A. Schneeberger, Duke University; Lawren M. Smithline, Harvard University; Robert G. Southworth, California Institute of Technology; Colin M. Springer, University of Waterloo; Jun Teng, California Institute of Technology; Andras Vasy, Stanford University; Martin M. Wattenberg, Brown University; David B. Wilson, Massachusetts Institute of Technology; Michael P. Wolf, Harvard University; and John H. Woo, Harvard University.

The other individuals who achieved ranks among the top 98, in alphabetical order of their schools, were: Boston University, Michael G. Szydlo; University of British Columbia, Gregory F. Wellman; University of Calgary, Geoffrey T. Falk; California Institute of Technology, Ian Agol, Alan I. Knutson, William M. Watson; University of California, Berkeley, Stephen P. Bard, Thomas S. Lumley, Max L. Shireson, Zheng Yin; Carleton University, Adam M. Logan; Carnegie Mellon University, Sanjay Khanna; University of Chicago, David J. Pollack, Adrian Tanner; Columbia University, Andrew Mogilyansky; Cornell University, Isaac J. Kuo; Harvard University, Daniel E. Gottesman, F. Dean Hildebrandt, Roger W. Lee, Andrew P. Lewis; Harvey Mudd College, Guy D. Moore; University of Illinois, Urbana-Champaign, David E. Bekman; University of Maryland, College Park, Lev Novik; Massachusetts Institute of Technology, Andrew Chou, Edward B. Hontz, Michael J. Lawler; Michigan State University, Thomas P. Hayes, Jacob R. Lorch; University of Minnesota, Twin Cities, Wei Shen; University of Missouri, Rolla, Xi Chen; Mount Allison University, Eugene Fink; New York University, Daniel J. Bernstein; State University of New York, Stony Brook, Jason Israel; Princeton University, Timothy Y. Chow, Gregory D. Landweber; Queens University, Alex Grossman; Reed College, Nathaniel J. Thurston; Stanford University, Beesham A. Seecharan, Jay A. Shrauner, Garrett R. Vargas; Swarthmore College, Olaf A. Holt; University of Texas, Austin, Bryan W. Taylor; University of Toronto, Nima Arkani-Hamed; Trinity College, Hartford, Marshall A. Whittlesey; Washington State University, Julie B. Kerr; Washington University, St. Louis, Peter H. Berman, Adam M. Costello, Jeremy T. Tyson; University of Waterloo, Michael A. Buckley, John Daniel Christensen; and Yale University, Thomas Zuwei Feng, Evan M. Gilbert, Zhaoliang Zhu.

There were 2347 individual contestants from 380 colleges and universities in Canada and the United States in the competition of December 1, 1990. Teams were entered by 289 institutions.

The Questions Committee for the fifty-first competition consisted of George E. Andrews, Paul R. Halmos (Chair), and Kenneth A. Stolarsky; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1. Let

$$T_0 = 2, \qquad T_1 = 3, \qquad T_2 = 6,$$

and for $n \geq 3$,

$$T_n = (n + 4)T_{n-1} - 4nT_{n-2} + (4n - 8)T_{n-3}.$$

The first few terms are

$$2, 3, 6, 14, 40, 152, 784, 5168, 40576, 363392.$$

Find, with proof, a formula for T_n of the form $T_n = A_n + B_n$, where (A_n) and (B_n) are well-known sequences.

Problem A-2. Is $\sqrt{2}$ the limit of a sequence of numbers of the form $\sqrt[3]{n} - \sqrt[3]{m}$, $(n, m = 0, 1, 2, \ldots)$?

Problem A-3. Prove that any convex pentagon whose vertices (no three of which are collinear) have integer coordinates must have area $\geq 5/2$.

Problem A-4. Consider a paper punch that can be centered at any point of the plane and that, when operated, removes from the plane precisely those points whose distance from the center is irrational. How many punches are needed to remove every point?

Problem A-5. If **A** and **B** are square matrices of the same size such that **ABAB** = **0**, does it follow that **BABA** = **0**?

Problem A-6. If X is a finite set, let $|X|$ denote the number of elements in X. Call an ordered pair (S, T) of subsets of $\{1, 2, \ldots, n\}$ *admissible* if $s > |T|$ for each $s \in S$, and $t > |S|$ for each $t \in T$. How many admissible ordered pairs of subsets of $\{1, 2, \ldots, 10\}$ are there? Prove your answer.

Problem B-1. Find all real-valued continuously differentiable functions f on the real line such that for all x

$$(f(x))^2 = \int_0^x \left((f(t))^2 + (f'(t))^2\right) dt + 1990.$$

Problem B-2. Prove that for $|x| < 1, |z| > 1$,

$$1 + \sum_{j=1}^{\infty} (1 + x^j) \frac{(1 - z)(1 - zx)(1 - zx^2) \cdots (1 - zx^{j-1})}{(z - x)(z - x^2)(z - x^3) \cdots (z - x^j)} = 0.$$

Problem B-3. Let S be a set of 2×2 integer matrices whose entries a_{ij} (1) are all squares of integers, and, (2) satisfy $a_{ij} \leq 200$. Show that if S has more than 50387 $(= 15^4 - 15^2 - 15 + 2)$ elements, then it has two elements that commute.

Problem B-4. Let G be a finite group of order n generated by a and b. Prove or disprove: there is a sequence

$$g_1, g_2, g_3, \ldots g_{2n}$$

such that

(1) every element of G occurs exactly twice, and

(2) g_{i+1} equals $g_i a$ or $g_i b$, for $i = 1, 2, \ldots, 2n$. (Interpret g_{2n+1} as g_1.)

Problem B-5. Is there an infinite sequence $a_0, a_1, a_2, \ldots$ of nonzero real numbers such that for $n = 1, 2, 3, \ldots$ the polynomial

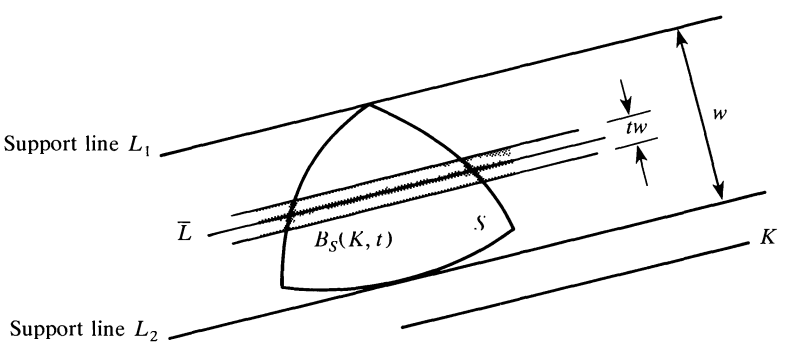
$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

has exactly n distinct real roots?

Problem B-6. Let S be a nonempty closed bounded convex set in the plane. Let K be a line and t a positive number. Let L_1 and L_2 be support lines for S parallel to K, and let $\bar{L}$ be the line parallel to K and midway between L_1 and L_2. Let $B_S(K, t)$ be the band of points whose distance from $\bar{L}$ is at most $(t/2)w$, where w is the distance between L_1 and L_2. What is the smallest t such that

$$S \cap \bigcap_K B_S(K, t) \neq \varnothing$$

for all S? (K runs over all lines in the plane.)



SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geq i \geq 0$ is the number of students among the top 199 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1 (150, 9, 1, 0, 0, 0, 0, 0, 1, 1, 6, 33)

Solution. The formula for T_n is

$$T_n = n! + 2^n.$$

This can be verified by induction. Alternatively, set $t_n = n! + 2^n$. Clearly $t_0 = 2 = T_0$, $t_1 = 3 = T_1$ and $t_2 = 6 = T_2$. Also,

$$t_n - nt_{n-1} = 2^n - n2^{n-1}.$$

Now 2^n and $n2^{n-1}$ are both solutions of the recurrence equation

$$f_n - 4f_{n-1} + 4f_{n-2} = 0, \qquad (*)$$

which is easily shown by direct substitution. Therefore since $t_n - nt_{n-1}$ is a linear combination of solutions to $(*)$, it must also be a solution. Consequently,

$$(t_n - nt_{n-1}) - 4(t_{n-1} - (n-1)t_{n-2}) + 4(t_{n-2} - (n-2)t_{n-3}) = 0,$$

or

$$t_n = (n+4)t_{n-1} - 4nt_{n-2} + (4n-8)t_{n-3}.$$

Hence $t_n = T_n$ because they are identical for $n = 0, 1, 2$ and satisfy the same third-order recurrence $(*)$ for $n \geq 3$.

A-2 (63, 25, 16, 4, 0, 0, 0, 4, 5, 6, 21, 57)

Solution. Since

$$\sqrt[3]{n+1} - \sqrt[3]{n} = \frac{1}{\sqrt[3]{(n+1)^2} + \sqrt[3]{(n+1)n} + \sqrt[3]{n^2}},$$

it follows that $\sqrt[3]{n+1} - \sqrt[3]{n} \to 0$ as $n \to \infty$, and hence that there are arbitrarily small numbers of the form $\sqrt[3]{n} - \sqrt[3]{m}$. Since $k(\sqrt[3]{n} - \sqrt[3]{m}) = \sqrt[3]{k^3 n} - \sqrt[3]{k^3 m}$, it follows that the set of numbers of that form is closed under multiplication by arbitrary positive integers. The preceding two sentences imply that the set of numbers of the form under consideration is dense, and hence that every real number is a limit of a sequence of such numbers.

A-3 (4, 4, 4, 0, 0, 0, 0, 0, 22, 36, 76, 55)

Solution. By Pick's formula, the area is $I + B/2 - 1$, where I is the number of internal lattice points and B is the number on the boundary. Clearly, $I \geq 0$ and $B \geq 5$. If $I \geq 1$ we are done.

If $I = 0$, then separate the vertices v_1, v_2, v_3, v_4, v_5 into four classes according to the parity of their coordinates. At least one class must have at least two elements, say v_1 and v_2. Hence the mid-point $\frac{1}{2}(v_1 + v_2)$, is also a lattice point; call it v_0. Since $I = 0$, v_0 is on the boundary of the pentagon. Now consider the five points $\{v_0, v_2, v_3, v_4, v_5\}$. The same reasoning produces a second lattice point v_0^* which is not v_1 (since v_0^* is a mid-point) and not in the interior (since $I = 0$.) Thus we have a second new lattice point on the boundary. Therefore, $B \geq 7$, so again the area is $\geq 5/2$.

A-4 (44, 7, 6, 6, 0, 0, 0, 0, 32, 15, 30, 61)

Solution. The answer is certainly greater than 2. Reason: to any two distinct points there corresponds at least one point whose distance from each of the given ones is rational. Proof: draw circles centered at the given points with rational radii; if the circles are not chosen too carelessly, they will intersect. (Choose the radii to be more than half the given distance but less than the whole.)

Three punches are enough. Indeed: punch twice, at distinct centers. Since each punch leaves countably many circles, the two punches leave their intersections, a countable set. Consider all circles centered at points of that set, with rational radii; their intersections with an arbitrary line form a countable set. A point of that line

not in that countable set is at an irrational distance from all remaining points; apply the punch there.

A-5 (29, 5, 0, 0, 0, 0, 0, 0, 1, 0, 58, 108)

Solution. The answer is yes for 2×2 matrices and no in all other cases. Indeed: since $\mathbf{ABAB} = \mathbf{0}$, it follows that $\mathbf{B(ABAB)A} = \mathbf{0}$, and hence that $\mathbf{BA}$ is nilpotent. If a 2×2 matrix $\mathbf{M}$ is nilpotent, then $\mathbf{M}^2 = 0$ (because the characteristic equation of $\mathbf{M}$ has degree 2 or less).

A counterexample for 3×3 (and therefore, just by enlargement by 0's, for any size) is to take

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A-6 (6, 6, 54, 1, 0, 0, 0, 0, 4, 0, 45, 85)

Solution. Let A_n denote the number of admissible ordered pairs of subsets of $\{1, 2, \ldots, n\}$. Clearly

$$A_n = \sum_{0 \le i, j \le n} \binom{n-i}{j}\binom{n-j}{i}.$$

Define

$$B_n = \sum_{0 \le i, j \le n} \binom{n+1-i}{j}\binom{n-j}{i}.$$

Clearly $A_0 = B_0 = 1$, and

$$\begin{aligned} B_n - A_n &= \sum_{0 \le i, j \le n} \binom{n-i}{j-1}\binom{n-j}{i} \\ &= \sum_{0 \le i, j \le n-1} \binom{n-i}{j}\binom{n-1-j}{i} \\ &= B_{n-1}, \end{aligned}$$

while

$$\begin{aligned} A_n - B_{n-1} &= \sum_{0 \le i, j \le n-1} \binom{n-i}{j}\binom{n-j-1}{i-1} + 2 \\ &= \sum_{0 \le i, j \le n-1} \binom{n-i-1}{j}\binom{n-j-1}{i} + 1 \\ &= A_{n-1} + 1. \end{aligned}$$

Hence we immediately verify by induction that

$$A_n = F_{2n+2}, \qquad B_n = F_{2n+3} - 1.$$

Hence, $A_{10} = F_{22} = 17711$. (F_i is the i-th Fibonacci number, defined by $F_0 = 0$, $F_1 = 1$, and for $n \ge 2$, $F_n = F_{n-1} + F_{n-2}$.)

B-1 (114, 2, 52, 0, 0, 0, 0, 11, 5, 3, 10, 4)

Solution. There are two such functions, namely $f(x) = \sqrt{1990}e^x$, and $f(x) = -\sqrt{1990}e^x$. To see this, suppose that the identity holds. Differentiating each side gives

$$2f(x)f'(x) = (f(x))^2 + (f'(x))^2,$$

or equivalently,

$$(f(x) - f'(x))^2 = 0, \qquad f'(x) = f(x),$$
$$\log|f(x)| = x + C, \qquad |f(x)| = e^C e^x.$$

But f is continuous and $f(0) = \pm\sqrt{1990}$, and this implies that $f(x) = \pm\sqrt{1990}e^x$.

B-2 (23, 5, 4, 9, 0, 0, 0, 3, 0, 0, 32, 125)

Solution. Let $S_0 = 1$, and for $n \geq 1$, let

$$S_n = 1 + \sum_{j=1}^{n} (1 + x^j)\frac{(1 - z)(1 - zx)(1 - zx^2)\cdots(1 - zx^{j-1})}{(z - x)(z - x^2)(z - x^3)\cdots(z - x^j)}.$$

It is easy to check that $S_1 = (1 - zx)/(z - x)$, $S_2 = (1 - zx)(1 - zx^2)/(z - x)(z - x^2)$, and by induction,

$$S_n = \frac{(1 - zx)(1 - zx^2)\cdots(1 - zx^n)}{(z - x)(z - x^2)\cdots(z - x^n)}.$$

To complete the proof, we need to prove that $\lim_{n\to\infty} S_n = 0$. To see this, we note that

$$S_{n+1} = \left(\frac{1 - zx^{n+1}}{z - x^{n+1}}\right)S_n.$$

As $n \to \infty$, $1 - zx^{n+1}$ goes to 1 and $z - x^{n+1}$ goes to z. Thus, there exist positive numbers N and ε such that

$$\left|\frac{1 - zx^{n+1}}{z - x^{n+1}}\right| < \frac{1}{|z|} + \varepsilon < 1$$

for all integers $n > N$. It follows that

$$|S_{n+1}| < \left(\frac{1}{|z|} + \varepsilon\right)|S_n|$$

and the result follows.

B-3 (97, 7, 4, 2, 0, 0, 0, 0, 12, 2, 54, 23)

Solution. Let U be all such 2×2 matrices, D the diagonal ones, and J those that are multiples of $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Note that (i) any two from D commute, (ii) any two from J commute, and (iii) $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 4 \\ 0 & 1 \end{smallmatrix}\right)$ commute. Clearly,

$$|U \cap (D \cup J)^c| = |U| - |D| - |J| + |D \cap J| = 15^4 - 15^2 - 15 + 1.$$

Suppose that no two elements of S commute, and write

$$S = (S \cap (D \cup J)) \cup (S \cap (D \cup J)^c).$$

Clearly, $|S \cap (D \cup J)| \leq 2$ and (by (iii) above)

$$|S \cap (D \cup J)^c| < |U \cap (D \cup J)^c| \leq 15^4 - 15^2 - 15 + 1.$$

The result follows. (Here X^c denotes the complement of X.)

The number 50387 is far from the best possible and there are many potential solutions. *Jiuqiang Liu* and *Allen J. Schwenk* from *Western Michigan University* have shown, using an inclusion-exclusion argument, that the maximum number of elements in U in which no two elements commute is 32390.

B-4 (9, 5, 2, 1, 0, 0, 0, 0, 3, 2, 63, 116)

Solution. Construct a graph whose vertices are labeled by the elements of G, so that for each vertex g, there are two "out" arcs, one to vertex ga and one to vertex gb (and consequently, each vertex g has two "in" arcs coming to it, one from ga^{-1} and one from gb^{-1}). The resulting graph is connected and each vertex has outdegree 2 and indegree 2. Therefore there is an Eulerian path which traverses the arcs, once and only once, and returns to the beginning. We get the desired sequence by listing the group elements associated with the vertices as we follow this path.

B-5 (16, 11, 15, 12, 0, 0, 0, 11, 5, 5, 48, 78)

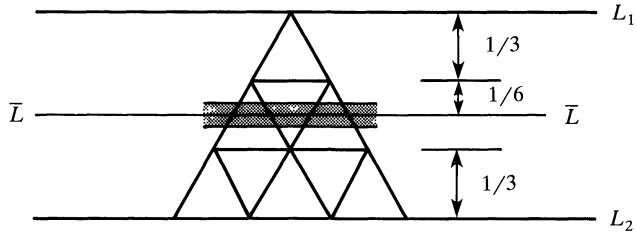
Solution. Take $a_0 = 1$, $a_1 = -1$, and proceed by induction. Say $p_n(x)$ has the property, and also $p_n(x) \to \infty$ or $-\infty$ as $x \to \infty$ depending upon whether n is even or odd. Then

$$p_{n+1}(x) \equiv p_n(x) + \frac{(-x)^{n+1}}{M}$$

has a sign change arbitrarily close to every root of $p_n(x)$ for M sufficiently large, and also the same sign as $p_n(x)$ at $x_n^* + 1$ where x_n^* is the largest root of $p_n(x)$. But now (M is already fixed) for x sufficiently large $p_{n+1}(x)$ has another sign change. Since $p_{n+1}(x)$ has at most $n + 1$ roots, the result follows.

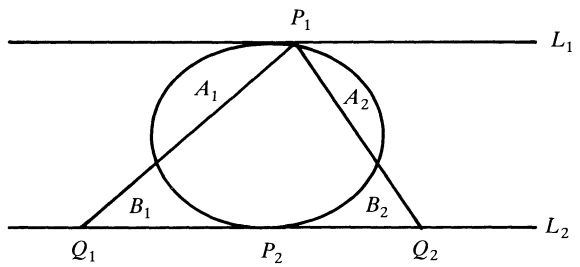
B-6 (5, 0, 0, 0, 0, 0, 0, 0, 38, 6, 29, 123)

Solution. Consider the dissection of the equilateral triangle (as shown) into 9 similar equilateral triangles.



It shows that any $t < 1/3$ produces an empty intersection.

We now show that the intersection is nonempty for $t \geq 1/3$, since it always contains the centroid of S. Think of L_1 as the upper support line (see sketch).



Let P_1 be a point of contact of L_1 with S, and P_2 a point of contact of L_2 with S. Extend two lines from P_1 to points Q_1, Q_2 on L_2 so that A_1, the region "above" P_1Q_1 and in S has area equal to that of B_1, the region "below" P_1Q_1 above L_2 and outside of S. If S is perturbed by replacing A_1 by B_1 (and similarly A_2 by B_2) the new S (it is a triangle) will have a "lower" centroid. But this new centroid is still $1/3$ of the way above L_2 (on the way to L_1). Hence if β is a band comprising the middle $1/3$ of the strip between L_1 and L_2, it contains the centroid of S. Hence for $t = 1/3$, the intersection is nonempty. Hence $t = 1/3$ is the smallest such value.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Branching Processes That Grow Faster Than Binary Splitting

F. M. DEKKING
Department of Mathematics and Information, Delft University of Technology, Julianalaan 132, 2628 BL Delft, The Netherlands[1]

A drastic qualitative change in a physical system occurring after an arbitrarily small change of a parameter (like temperature) is usually called a critical phenomenon. One of the most thoroughly studied mathematical models of such phenomena is that of percolation which paraphrases the passage of a fluid through a porous medium consisting of very many small channels ([2]). The qualitative property of interest is whether or not the fluid will percolate the medium macroscopically if it passes each small channel with probability p (the parameter of the process). Typically this has probability zero for all values of p smaller than a critical value p_c, and strictly positive probability for p larger than p_c.

For most percolation models it is very difficult to determine the exact value of p_c, and even more difficult to describe the behavior of the probability of percolation near p_c. These difficulties are mainly caused by the interconnectivity structure of the small channels in the medium. In this note we present a model which has the same flavor as percolation models (where the interconnectivity structure is that of a tree), in which the above mentioned quantities can be exactly computed in an elementary way. The relation to existing models is in fact very close: our model has been derived from the analysis of Mandelbrot's percolation process in [1].

1. Branching processes and family trees. The theory of branching processes is often said to have its birth with the solution by H. W. Watson of the problem of "the decay of the families of men" posed by F. Galton (see, e.g. [3]). This problem can be phrased as follows. We are given an offspring distribution Z_1, i.e. a random variable taking non-negative integer values k with probability $p_k \geqslant 0$, $k = 0, 1, \ldots (\sum_{k=0}^{\infty} p_k = 1)$. At time 0, there is one particle, the ancestor ($Z_0 \equiv 1$). This particle gives rise to a certain number of particles at time 1, according to the distribution Z_1. Each of the Z_1 particles at time 1 independently gives rise to a certain number of particles at time 2, each according to the offspring distribution Z_1. At time n there are Z_n particles each reproducing independently of all particles at the same or previous time instants, all according to the same offspring distribution Z_1, giving rise to a total of Z_{n+1} particles at time $n + 1$. Here it is understood that $Z_{n+1} \equiv 0$ if $Z_n \equiv 0$. The Galton-Watson problem is to determine the probability that the process (Z_n) *becomes extinct*, i.e. that $Z_n \equiv 0$, for some $n \geqslant 1$.

[1] Work performed during a stay at the Université de Provence, Marseille, France

With a branching process (Z_n) we may associate a (random) family tree: there is a distinguished node called the *root* associated to the ancestor, which is connected to Z_1 first level nodes. In general there are Z_n n^{th} level nodes, and each n^{th} level node is connected to the $n + 1^{\text{th}}$ level nodes associated with the offspring of the

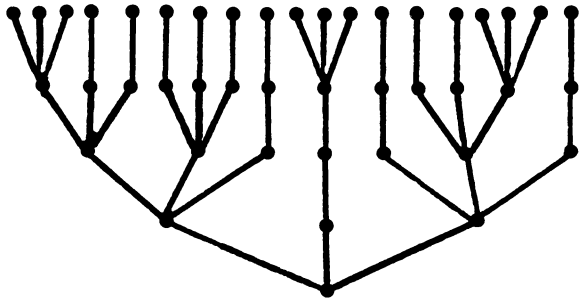


FIG. 1

corresponding particle at time n. As an example, consider the (1-or-3)-process defined by $p_1 = 1 - p$ and $p_3 = p$ for some p between 0 and 1. See Figure 1 for a realization of the first four levels of the family tree of such a process. Note that the family tree associated to a branching process is finite if and only if the process becomes extinct, and that the extinction problem can now be interpreted as the question "with what probability does the family tree contain the unary tree as a subtree, rooted at the root of the family tree?" Here the *unary tree* is the family tree associated to the (deterministic) branching process where each particle gives rise to exactly one particle (the offspring distribution is determined by $p_1 = 1$).

We shall be concerned with the question "with what probability does the family tree of a branching process contain the binary tree as a subtree, rooted at the root of the family tree?" Here the *binary tree* is the family tree associated to the branching process with offspring distribution given by $p_2 = 1$ (each particle gives rise to exactly two particles). If this probability is positive, then we say that the branching process *grows faster than binary splitting*, or has the *binary splitting property*.

2. A fixed point equation. The solution of the extinction problem makes use of the *generating function* f of the offspring distribution defined by

$$f(x) = \sum_{k=0}^{\infty} p_k x^k.$$

In fact, let δ_n be the probability that $Z_n \equiv 0$. Then one finds, intersecting the event $[Z_{n+1} = 0]$ with the events $[Z_1 = k]$ that

$$\delta_{n+1} = p_0 + \sum_{k=1}^{\infty} \delta_n^k p_k = f(\delta_n).$$

Hence $\delta = \lim_{n\to\infty} \delta_n$ satisfies the equation

$$f(x) = x. \tag{2.1}$$

Putting $\delta_0 = 0$, one sees that δ is the smallest root in $[0,1]$ of (2.1). Then (excluding the trivial case where $p_1 = 1$, in which case $\delta = 0$) using that f is a convex function with $f(1) = 1$ and $f'(1) = EZ_1$, the mean of the offspring distribution, one proves the well-known extinction criterion for a nontrivial branching process (Z_n):

$$\delta = 1 \quad \text{iff} \quad EZ_1 \leqslant 1. \tag{2.2}$$

As we will see later, there is no hope for such a simple criterion for the binary splitting property. First we prove an analogue of (2.1).

THEOREM 1. *Let (Z_n) be a branching process with generating function f, and let π be the probability that (Z_n) grows faster than binary splitting. Then $1 - \pi$ is the smallest root in* $[0,1]$ *of the equation*

$$f(x) + (1-x)f'(x) = x. \tag{2.3}$$

Proof. For $n = 1, 2, \ldots$ let $\gamma_n = 1 - \pi_n$ be the probability that the family tree associated to (Z_n) does *not* contain a binary subtree of height n rooted at the root of the family tree. If the family tree has k nodes at level 1 (i.e., $Z_1 = k$), then it does not contain a binary subtree of height $n+1$ iff $k = 0$ or 1, or either all or all but one of the k subtrees rooted at these nodes do not contain a binary subtree of height n. Therefore one has

$$\begin{aligned}\gamma_{n+1} &= p_0 + p_1 + \sum_{k=2}^{\infty} \left(\gamma_n^k + k\gamma_n^{k-1}(1-\gamma_n)\right)p_k \\ &= p_0 + p_1 + f(\gamma_n) - p_0 - p_1\gamma_n + (1-\gamma_n)\left(f'(\gamma_n) - p_1\right),\end{aligned}$$

and hence

$$\gamma_{n+1} = G(\gamma_n), \tag{2.4}$$

where $G(x) = f(x) + (1-x)f'(x)$. Therefore $\lim_{n\to\infty} \gamma_n =: \gamma = 1 - \pi$ satisfies (2.3). Furthermore, if we put $\gamma_0 = 0$, then (2.4) is also true for $n = 0$. Since $G'(x) = (1-x)f''(x)$ is non-negative, $G(x)$ is increasing and it follows that $\gamma = \lim_{n\to\infty} G^n(0)$ is the *smallest* root in $[0,1]$ of (2.3). ■

A first simple consequence of this result (which is also obvious directly) is that $\pi = 1$ if and only if $p_0 + p_1 = 0$. There is no simple expression involving the moments of the offspring distribution which yields whether $\pi = 0$ or not. We shall return to this remark in the next section, giving now an example to illustrate this. Let (Z_n) be any branching process with offspring distribution given by $(p_k)_{k=0}^{\infty}$, and let (Z_n') be the branching process with offspring distribution $(p_k')_{k=0}^{\infty}$ given by

$$p_0' = 0, \qquad p_1' = p_0 + p_1, \qquad p_k' = p_k \qquad \text{for } k \geqslant 2.$$

Then it is clear (if not by comparing the associated family trees, then because they yield the same equation (2.3)), that π is the same for both processes. However, if $p_0 > 0$, then all moments of the first offspring distribution are strictly smaller than those of the second.

3. A critical phenomenon. Here we shall complement the remark made in the previous section by giving an example which shows that the binary splitting probability π can not be a continuous function of the moments of the offspring distribution, nor of any other parameter of the process which depends continuously on the p_k, $k = 0, 1, 2, \ldots$.

THEOREM 2. *For each p in* $[0, 1]$ *let* (Z_n) *be the branching process determined by* $p_1 = 1 - p$, $p_3 = p$, *and let* $\pi(p)$ *be the binary splitting probability of* (Z_n). *Then*

$$\pi(p) = \begin{cases} 0 & \text{for } 0 \leqslant p < \dfrac{8}{9} \\ \dfrac{3}{4} + \dfrac{3}{4}\sqrt{1 - \dfrac{8}{9p}} & \text{for } \dfrac{8}{9} \leqslant p \leqslant 1. \end{cases}$$

Proof. The generating function of the offspring distribution is $f(x) = (1 - p)x + px^3$. According to Theorem 1, $1 - \pi(p)$ is the smallest root in $[0, 1]$ of

$$1 - p + 3px^2 - 2px^3 = x, \tag{3.1}$$

which is equivalent to

$$1 - x = p(2x^3 - 3x^2 + 1).$$

Dividing by $(1 - x)$ we obtain $1 = -p(2x^2 - x - 1)$, which has roots given by

$$x_{1,2} = \frac{1}{4} \pm \frac{1}{4p}\sqrt{p(9p - 8)}\,.$$

Hence, if $p < \frac{8}{9}$ then the smallest root of (3.1) is necessarily $x = 1$, and therefore $\pi(p) = 0$; if $p \geqslant \frac{8}{9}$ then the smallest root of (3.1) is

$$\frac{1}{4} - \frac{1}{4p}\sqrt{p(9p - 8)}\,, \text{ so } \pi(p) = \frac{3}{4} + \frac{1}{4}\sqrt{9 - \frac{8}{p}}\,. \quad \blacksquare$$

We end this note by remarking that Theorem 1 and 2 can obviously be generalized. It is an easy exercise to compute $\pi(p)$ for the branching process with geometric offspring distribution ($p_k = (1 - p)p^k$ for $k \geqslant 0$). The critical value of p here is $\frac{4}{5}$. See [**4**] for generalizations to N-ary subtrees.

REFERENCES

1. J. T. Chayes, L. Chayes, and R. Durrett, Connectivity properties of Mandelbrot's percolation process. *Probab. Th. Rel. Fields*, 77 (1988) 307–324.
2. G. R. Grimmett, Percolation, Springer Verlag, Berlin, 1989.
3. T. E. Harris, The Theory of Branching Processes, Springer Verlag, Berlin, 1963.
4. A. G. Pakes and F. M. Dekking, On family trees and subtrees of simple branching processes, to appear in J. Theoret. Probab. 4 (1991).

by

$$g_{ij}^{*} = \begin{cases} D^{j-i}[g]/(j-i)! & \text{for } 1 \leqslant i \leqslant j \leqslant n+1 \\ 0 & \text{for } 1 \leqslant j < i \leqslant n+1 \end{cases}$$

The article [2] referring to the approach in [1] includes the following:

THEOREM. *The map $g \to g^*$ is a homomorphism from the algebra of functions continuously differentiable on a neighborhood N to a commutative algebra of upper-triangular matrices.*

Hence we have the following statements:

(1) $(f+g)^* = f^* + g^*$,

(2) $(cf)^* = c \cdot f^*$, where c is a real constant,

(3) $(f \cdot g)^* = f^* \cdot g^* = g^* \cdot f^*$,

(4) $(1/g)^* = (g^*)^{-1}$ if $g(x) \neq 0$,

(5) $(f/g)^* = f^* \cdot (g^*)^{-1} = (g^*)^{-1} \cdot f^*$ if $g(x) \neq 0$.

A similar result for the map $g \to G_g$ was discovered independently in the process of this research but can be nicely proved from the above theorem by noticing that $G_g = (Dg^*D^{-1})^t$, where $D = \operatorname{diag}\{0!, 1/1!, 1/2!, \ldots, 1/n!\}$. For example, a proof of the statement comparable to (4) above would be the following:

$$G_{1/g} = \left[D(1/g)^*D^{-1}\right]^t$$

$$= \left[D(g^*)^{-1}D^{-1}\right]^t = \left[\left(Dg^*D^{-1}\right)^{-1}\right]^t = \left[\left(Dg^*D^{-1}\right)^t\right]^{-1} = G_g^{-1}.$$

As a matter of fact, any mapping of the form $g \to Vg^*V^{-1}$ or $g \to (Vg^*V^{-1})^t$, where V is a constant, nonsingular matrix, is an algebraic homomorphism as above.

Thus we have,

$$G_g^{-1} = G_{1/g} = \begin{bmatrix} 1/g & 0 & & & \cdots & 0 \\ D[1/g] & 1/g & 0 & & \cdots & 0 \\ D^2[1/g] & 2D[1/g] & 1/g & 0 & \cdots & 0 \\ D^3[1/g] & 3D^2[1/g] & 3D[1/g] & 1/g & \cdots & 0 \\ D^n[1/g] & nD^{n-1}[1/g] & & & \cdots & 1/g \end{bmatrix}$$

A second approach to G_g^{-1} is by way of the Cayley-Hamilton Theorem of linear algebra. Here we have the obvious extension of the theorem in view of the fact that G_g is a matrix of functions defined on some real neighborhood.

THEOREM (*Cayley-Hamilton*). *For each value x in the domain of g, the matrix G_g satisfies its characteristic equation for that value of x.*

But for a given value of x the characteristic equation of G_g has the form

$$[g(x) - \lambda(x)]^{n+1} = 0,$$

where $\lambda(x)$ represents an eigenvalue of G_g for this specific value of x.

Thus, for each x, we have the matrix equation

$$[g \cdot I - G_g]^{n+1} = 0.$$

Using the binomial theorem, one can expand this expression to obtain

$$\sum_{k=0}^{n+1} \binom{n+1}{k} g^{n+1-k}(-1)^k G_g^k = 0.$$

where it is understood that $G_g^0 = I$. With some algebraic manipulation one obtains

$$G_g \cdot \sum_{k=1}^{n+1} \binom{n+1}{k} g^{n+1-k}(-1)^k G_g^{k-1} = -g^{n+1} \cdot I.$$

Hence we have the formula

$$G_g^{-1} = \sum_{k=0}^{n} (-1)^k \binom{n+1}{k+1} \frac{1}{g^{k+1}} G_g^k.$$

But thankfully we can write G_g^k as G_{g^k} by the homomorphism, so that the solution to the earlier system of equations represented as $G_g \cdot X = E$ is

$$X = \sum_{k=0}^{n} (-1)^k \binom{n+1}{k+1} \frac{1}{g^{k+1}} G_{g^k} \cdot E.$$

The last component of this vector is $D^n[1/g]$, from the prior approach, so that we have the following result:

THEOREM.

$$D^n[1/g] = \sum_{k=1}^{n} (-1)^k \binom{n+1}{k+1} \frac{1}{g^{k+1}} D^n[g^k].$$

[Note that the sum starts with $k = 1$ since for $k = 0$ and $n \geqslant 1$, $D^n[g^k] = 0$.]

With this formula in hand one can try some applications. Choose your favorite function, g, and try it! The problem, of course, is that the formula requires the nth derivatives of g^k for $k = 1, 2, \ldots, n$. Nevertheless, there are a few cases in which it is easy and some interesting *combinatorial* results arise!

With $g(x) = e^x$ one obtains:

$$(-1)^n = \sum_{k=1}^{n} (-1)^k \binom{n+1}{k+1} k^n.$$

With $g(x) = x^m$, for any real number m, one obtains

$$(-1)^n (m+n-1) \cdots (m+1)m = \sum_{k=1}^{n} (-1)^k \binom{n+1}{k+1} (mk)(mk-1) \cdots (mk-n+1).$$

If m is a positive integer, then dividing by $n!$ yields the identity:

$$(-1)^n \binom{m+n-1}{n} = \sum_{k=1}^{n} (-1)^k \binom{n+1}{k+1} \binom{mk}{n}$$

If m is a negative integer we note the following:

$$(m+n-1)\cdots(m+1)m = \begin{cases} 0 & \text{for } n > -m \\ (-1)^n n!\binom{-m}{n} & \text{for } n \leqslant -m. \end{cases}$$

Thus we have the two additional cases. If $n > -m$ then

$$0 = \sum_{k=1}^{n} (-1)^{n+k}\binom{n+1}{k+1}\binom{n-mk-1}{n},$$

whereas if $n \leqslant -m$ then

$$\binom{-m}{n} = \sum_{k=1}^{n} (-1)^{n+k}\binom{n+1}{k+1}\binom{n-mk-1}{n}.$$

REFERENCES

1. F. R. Gantmacher, The Theory of Matrices, Vol. 1, Chelsea Publishing Company, New York, 1959.
2. Yasuhiko Ikebe and Toshiyuki Inagaki, An elementary approach to the functional calculus for matrices, this MONTHLY, 93 (1986) 390–392.

Nested Radicals

JONATHAN M. BORWEIN[1]
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada

G. DE BARRA
Department of Mathematics, R.H.B.N.C., University of London, Egham, TW20 0EX, UK

1. Introduction. Ramanujan introduced the nested radicals

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\cdots}}} \tag{1}$$

and

$$\sqrt{6+2\sqrt{7+3\sqrt{8+4\cdots}}} \tag{2}$$

in [**1**] (see also [**4**, p. 327]) and gave solutions without any rigorous proof. It is not obvious that such expressions have an unambiguous meaning. In [**3**, p. 32] nested radicals of the form

$$\sqrt{a+\sqrt{a+\sqrt{a+\cdots}}} \tag{3}$$

are considered and convergence is shown to a root of the appropriate quadratic equation. A more difficult example with alternating signs is considered in [**5**, p. 299]. A solution had been given by Ramanujan and is to be found in

[1]Research supported by NSERC.

If m is a negative integer we note the following:

$$(m+n-1)\cdots(m+1)m = \begin{cases} 0 & \text{for } n > -m \\ (-1)^n n!\binom{-m}{n} & \text{for } n \leqslant -m. \end{cases}$$

Thus we have the two additional cases. If $n > -m$ then

$$0 = \sum_{k=1}^{n} (-1)^{n+k}\binom{n+1}{k+1}\binom{n-mk-1}{n},$$

whereas if $n \leqslant -m$ then

$$\binom{-m}{n} = \sum_{k=1}^{n} (-1)^{n+k}\binom{n+1}{k+1}\binom{n-mk-1}{n}.$$

REFERENCES

1. F. R. Gantmacher, The Theory of Matrices, Vol. 1, Chelsea Publishing Company, New York, 1959.
2. Yasuhiko Ikebe and Toshiyuki Inagaki, An elementary approach to the functional calculus for matrices, this MONTHLY, 93 (1986) 390–392.

Nested Radicals

JONATHAN M. BORWEIN[1]
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada

G. DE BARRA
Department of Mathematics, R.H.B.N.C., University of London, Egham, TW20 0EX, UK

1. Introduction. Ramanujan introduced the nested radicals

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\cdots}}} \tag{1}$$

and

$$\sqrt{6+2\sqrt{7+3\sqrt{8+4\cdots}}} \tag{2}$$

in [**1**] (see also [**4**, p. 327]) and gave solutions without any rigorous proof. It is not obvious that such expressions have an unambiguous meaning. In [**3**, p. 32] nested radicals of the form

$$\sqrt{a+\sqrt{a+\sqrt{a+\cdots}}} \tag{3}$$

are considered and convergence is shown to a root of the appropriate quadratic equation. A more difficult example with alternating signs is considered in [**5**, p. 299]. A solution had been given by Ramanujan and is to be found in

[1]Research supported by NSERC.

on taking $d := k^{\alpha/(p-1)}$. Set $\Delta(k) := |f_1(k) - f_2(k)|$. By (4) and (7)

$$\Delta(k) \leq \Delta(k + a)/p$$

so that

$$0 \leq \Delta(k) \leq \Delta(k + na)/p^n.$$

Thus any two solutions satisfying (6) agree for k. Hence by (4) they agree for all points of the form $k + \mathbb{N}a$. ■

COROLLARY. *Suppose that f satisfies* (4) *and* (6) *for all positive k and that a solution f^* to* (4) *is obtained by iteration as above commencing with $0 \leq f_0 \leq f$. Then $f^* = f$.*

Proof. From (5) we have that $f_n \leq f$ implies $f_{n+1} \leq f$, so $f^* \leq f$ and satisfies (6). By the Theorem $f^* = f$. ■

3. Examples. I. Take $p := 2$, $\alpha := 1$ and $g(k) := (\lambda - a)k + \lambda^2$ $(\lambda \geq a)$. Then the resulting functional equation

$$f(k) = \sqrt{(\lambda - a)k + \lambda^2 + kf(k + a)}$$

has solution $f(k) = k + \lambda$.

Further specialization to $\lambda = a$ produces a family of nested radicals including (1) which comes on setting $a = 1$. That is

$$f(2) = \sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + \cdots}}}$$

and so the value of (1) is 3, on using the Corollary and on defining the value of a nested radical such as (1) or (2) to be that provided by the iteration (5).

Similarly letting $\lambda = a + 1$ produces the functional equation

$$f(k) = \sqrt{k + (a + 1)^2 + kf(k + a)}$$

with solution $f(k) = k + a + 1$. Then $a := 1$, $k := 2$ gives the value of 4 for the expression (2), agreeing with Ramanujan.

II. Take $p := 3$, $\alpha := 2$, $a := 2b > 0$, and $g(k) := b^3 + 3b^2k$. Then (4) becomes

$$f(k) = \sqrt[3]{b^3 + 3b^2k + k^2 f(k + 2b)}$$

with solution $f(k) := k + b$. So iterating from any f_0 with $0 \leq f_0(k) \leq (b^3 + 3b^2k)^{1/3}$, $f_n(k)$ will converge to $k + b$. In particular $k := b := 1$ gives

$$2 = \sqrt[3]{4 + \sqrt[3]{10 + 9\sqrt[3]{16 + 25\sqrt[3]{22 + 49\sqrt[3]{28 + \cdots}}}}}. \tag{8}$$

III. Similarly, expansion of $(k + a)^4$ for $a > 0$ gives the solution to the functional equation

$$f(k) = \sqrt[4]{g(k) + k^3 f(k + 3a)},$$

where $g(k) := a^4 + 6a^2k^2 + 4a^3k$, as $f(k) := k + a$ and expressions similar to (8) follow on iteration from $f_0 := a$.

IV. In (5) take, for $k \geq 1$ (by translation), $\alpha := 0$, $p := z$, $a := 1$ and $g(k) := k$. Starting with $f_0 := 1$ we see easily that $f_n(k) \leq k + 1$ for all $k \geq 1$ and that $f_n(k)$

converges to the solution of $f(k) = \sqrt{(k + f(k+1))}$. Then

$$f(1) = \sqrt{1 + \sqrt{2 + \sqrt{3 + \cdots}}}\,.$$

This is the kind of "continued square root" considered in [7].

V. In (4) take $\alpha := 0$, $a := 0$ and $g(k) := k$, for $p > 1$ to get $f(k) = (k + f(k))^{1/p}$. This is solved by the iteration $f_0 := 1$ and $f_{n+1}(k) := (k + f_n(k))^{1/p}$. Then $f_n(k) \le (p-1)k + 1$ holds inductively. Consequently we have convergence to

$$\beta = \sqrt[p]{k + \sqrt[p]{k + \sqrt[p]{k + \cdots}}}$$

and β is finite. It follows that β is the positive root of $x^p - x - k = 0$.

4. A two variable extension. For simplicity we consider only square roots and wish to solve the functional equation

$$f(x,y) = \sqrt{g(x) + y^\alpha f(x + c, y + d)} \tag{9}$$

for a nonnegative function f where $x, y \ge 1$, $c, d, \alpha \ge 0$, g is a nonnegative function and $\gamma := \inf\{g(x)\colon x \ge 1\} > 0$. The results here are less satisfactory as no explicit solution of (9) is at hand. Again, we solve iteratively with $f_0 := 1$, and

$$f_{n+1}(x,y) = \sqrt{g(x) + y^\alpha f_n(x + c, y + d)}\,.$$

Whence

$$f_n^2(x,y) - f_{n-1}^2(x,y) = y^\alpha\big[f_{n-1}(x + c, y + d) - f_{n-2}(x + c, y + d)\big] \tag{10}$$

and $f_1(x,y) = \sqrt{(g(x) + y^\alpha)} \ge 1 = f_0(x,y)$. So, by (10) $f_n \ge f_{n-1}$ for all n and f_n increases to some f^* satisfying (9) and possibly taking infinite values.

UNIQUENESS. Suppose f and h satisfy (9). Then

$$f^2(x,y) - h^2(x,y) = y^\alpha\big[f(x + c, y + d) - h(x + c, y + d)\big].$$

So if $\Delta(x,y) := |f(x,y) - h(x,y)|$, it follows that

$$\Delta(x,y) = \frac{y^\alpha}{h(x,y) + f(x,y)}\Delta(x + c, y + d).$$

As before, we obtain from (9) for f (or h), $f(x,y) \ge \gamma^{1/2}$ and substituting back in (9) we obtain in the limit $f(x,y) \ge y^\alpha$. Thus

$$\Delta(x,y) \le 2^{-n}\Delta(x + nc, y + nd)$$

and, assuming a growth condition on f (and h): $f(x + nc, y + nd) = o(2^n)$, we get $f = h$.

EXAMPLE. Set $g(x) := (\lambda - a)x + \lambda^2 (\lambda \ge a)$, $c := d := a$, $\alpha := 1$. We may demand that $y := x$ and obtain Example I above.

5. Other solutions. The functional equation (4) can give rise to solutions which do not meet the growth condition (6). Consider (4) with $p := 2$, $g(k) := 1$, $a = \alpha := 1$. Then we have the simple functional equation

$$f(k) = \sqrt{1 + kf(k+1)}\,. \tag{11}$$

Let $g(k)$ be any non-trivial nonnegative C^∞ function with support in $[.25, .75]$ so that g and all its derivatives vanish at those endpoints. Extend g to be 0 elsewhere. Then let $f(k) := k + 1 + g(k)$ on $[0, 1)$ and use (11) to extend f to $[0, \infty)$. Then f is a nonnegative C^∞ solution to (11) but badly violates (6).

Similarly, consider the function f given by $f(k) = k + 1$ for $k \neq 1, 2, 3 \ldots$, and set $f(1) := 2 + t$, $t > 0$. Then (11) determines f on the remaining natural numbers and we have produced a finite discontinuous solution to (11). Moreover, from (11) we easily obtain, for an integer $m > 1$,

$$f(m) > m + 1 + 2^m t,$$

which violates the growth condition (6).

It is reasonable to conjecture that any analytic solution to (11) must satisfy (6) and so be $k + 1$. Anecdotal evidence for this is provided by the fact that the corresponding Riccati differential equation

$$x\frac{dy}{dx} = y^2 - xy - 1$$

has solutions $y_1 := (x + 1)$ and $y_2 := (x + 1) + x^2(1 - x - Ae^{-x})^{-1}$.

REFERENCES

1. Question 289, *Journal of the Indian Math. Soc. III* (1911) 90.
2. Notebooks of Srinivasa Ramanujan, Tata Institute, 1957.
3. *Problem* 460, *Amer. Math. Monthly*, 24 (1917)
4. G. H. Hardy, P. V. Seshu Aiyar, and B. N. Wilson Editors, *Collected Papers of Srinivasa Ramanujan*, *Cambridge University Press*, 1927.
5. Problem 1174 *Mathematics Magazine*, 57 (1984).
6. Putnam Competition, problem A6, *Amer. Math. Monthly*, 74 (1967).
7. W. S. Sizer, Continued roots, *Math. Magazine*, 59 (1986) 23–27.

Acknowledgement. We would like to thank John Giles and the Department of Mathematics at the University of Newcastle, Australia, and the Centre for Mathematical Analysis at the Australian National University for their support and hospitality while this work was in progress in 1988.

On Infinite and Finite Covering Systems[1]

M. A. BERGER, A. FELZENBAUM, A. S. FRAENKEL, AND R. HOLZMAN
Faculty of Mathematical Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

The integers and positive integers are denoted by $\mathbb{Z}$ and $\mathbb{Z}^+$, respectively. By $b \pmod a$ we denote the arithmetic sequence of integers $\{na + b: n \in \mathbb{Z}\}$, where $a \in \mathbb{Z}^+$ is the *modulus* and b, with $0 \leq b < a$, is the *residue* of a.

A system $\mathcal{S} = \{b_i \pmod{a_i}: i \in I\}$ of arithmetic sequences that forms a partition of $\mathbb{Z}$ is called a *disjoint covering system* (abbreviated DCS). Here $I =$

[1]This research was supported in part by Grant No. 85-00368 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. The work of A. S. Fraenkel was done during a visit at the Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada, during the summer of 1988. Support from Richard Guy's Canadian Research Grant 69-0695 is gratefully acknowledged.

$\{1, \ldots, N\}$ is a finite index set, or $I = \{1, 2, \ldots\}$ is countably infinite; $\mathscr{S}$ is said to be *finite* or *infinite* accordingly. A system $\mathscr{S}$ is called *incongruent* if $i \neq j \Rightarrow a_i \neq a_j$. It is well known that a DCS with $1 < |I| < \infty$ cannot be incongruent. See e.g. [8]. Thus, a finite DCS will always have at least two sequences with the same modulus (unless $|I| = 1$). Hence we shall speak about the *multiset* of moduli; the term multiset generalizes that of a set by allowing the membership of a specified number of copies of each element.

A major open problem in the study of DCS's is: given a multi-set of moduli, determine if it is possible to choose corresponding residues so as to obtain a DCS. Considering the finite case, Znám [8] gave a necessary condition on the multi-set of moduli for this to be possible, and conjectured that his condition was also sufficient. Recently Simpson [5] found a counterexample to Znám's conjecture, and gave a stronger necessary condition, in the hope that it might be sufficient; Zeilberger [7] frustrated this hope as well.

Attempts have also been made to characterize infinite incongruent DCS's (abbreviated IIDCS's). Early work on this usually appeared together with work on finite DCS's, such as in Stein [6], who asked whether the moduli of every nontrivial IIDCS are necessarily all powers of 2, and Krukenberg [4] who answered this in the negative. See also Dewar [2] and the survey article by Erdös [3]. Recently Beebee [1] raised renewed interest in the subject by asking about the existence of an IIDCS whose set of moduli was specified or was required to have certain properties.

This note is a contribution towards the unification of these two directions of research. We strengthen Znám's necessary condition while at the same time extending it to apply to infinite DCS's as well. As an application, we give a negative answer to one of Beebee's questions and to two similar questions.

The concept of the density d of a subset of $\mathscr{Z}$ is central to our argument, so we start by collecting a few well-known or easy facts about it. A subset $S \subseteq \mathscr{Z}$ is *periodic* if there exists $t \in \mathscr{Z}^+$, called a *period* of S, such that $S + t = S$. The *density* of a periodic set S is defined as $d(S) = |S \cap B_t|/t$, where t is any period of S and B_t is any block of t consecutive integers. This definition is consistent with the more general definition of asymptotic density, which we do not require here. It is easy to verify that the family $\mathscr{P}$ of periodic subsets of $\mathscr{Z}$ is closed under finite unions, and $d(\bigcup_{i=1}^{N} S_i) = \sum_{i=1}^{N} d(S_i)$ whenever $S_1, \ldots, S_N$ are periodic and pairwise disjoint. This property, called *finite additivity*, does not extend to infinite unions: First, $\bigcup_{i=1}^{\infty} S_i$ need not be periodic even though each of the S_i is periodic; second, even if $S_i \in \mathscr{P}$ $(i = 1, 2, \ldots)$ are pairwise disjoint and $\bigcup_{i=1}^{\infty} S_i \in \mathscr{P}$, it need not be the case that $d(\bigcup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} d(S_i)$. What is true, however, under these assumptions is that $d(\bigcup_{i=1}^{\infty} S_i) \geq \sum_{i=1}^{\infty} d(S_i)$; this is seen by observing that $d(\bigcup_{i=1}^{\infty} S_i) \geq d(\bigcup_{i=1}^{N} S_i) = \sum_{i=1}^{N} d(S_i)$, and letting $N \to \infty$. We shall refer to this property of d as *superadditivity* of the density.

Clearly $d(b \pmod a) = 1/a$, which suggests that the reciprocals of the moduli of a DCS must satisfy certain conditions. For a multiset of positive integers A, we denote $r(A) = \sum_{a \in A} 1/a$. Superadditivity of the density implies that if A is the multiset of moduli of a DCS then $r(A) \leq 1$; we say that the DCS is *saturated* if $r(A) = 1$. Examples of both saturated and unsaturated IIDCS's were given in Stein [6]. We restrict attention here to saturated DCS's, because unsaturated ones

are, in a sense, too easy to obtain. Finite additivity of the density guarantees that every finite DCS is saturated.

As usual, denote by (a, a') the greatest common divisor of a and a'.

THEOREM. *Let A be the multiset of moduli of a saturated DCS. Then for every prime p there exists a partition of A into multisets $A_0, A_1, \ldots, A_p$ such that*:

(i) $a \in A_p$ *if and only if* $p \nmid a$.

(ii) $r(A_i) = (1 - r(A_p))/p$ *for* $i = 0, \ldots, p - 1$.

(iii) *For each* $i \in \{0, \ldots, p - 1\}$ *with* $|A_i| > 1$, *if* $a, a' \in A_i$ *then* $(a, a') > p$.

Proof. Define the partition of A as follows. For every sequence $b \pmod a$ in the DCS with $p|a$, put a copy of a into that A_i, $i \in \{0, \ldots, p - 1\}$, for which $b \in i \pmod p$; put those $a \in A$ with $p \nmid a$ into A_p. Observe that each sequence $b \pmod a$ with $a \in A_i$, $i < p$, is contained in $i \pmod p$. If two such sequences (with the same i) had $(a, a') = p$ they would have to intersect; hence (iii) holds. Next, observe that each sequence with $a \in A_p$ is split by the residue classes $(\text{mod } p)$ into p sequences with modulus pa each. Hence, the superadditivity of the density, applied to the partition of the periodic set $i \pmod p$ induced by the DCS, yields

$$r(A_i) + \frac{1}{p} r(A_p) \leq \frac{1}{p}, \qquad i = 0, \ldots, p - 1.$$

Summing these p inequalities gives $r(A) \leq 1$. By assumption $r(A) = 1$, so each of the p weak inequalities must actually be an equality. This establishes (ii). ∎

REMARKS.

1. Znám's necessary condition, stated for finite DCS's, was the existence, for every prime p, of a partition of A into $A_0, A_1, \ldots, A_p$ satisfying (i) and (ii).
2. Our necessary condition is not sufficient; examples in Simpson [**5**] or Zeilberger [**7**] can be used to show this.
3. Denote $A(p) = \{a \in A: p|a\}$. It is a corollary of the Theorem that if $a \in A(p)$ then $r(A(p)) \geq p/a$. Indeed, let $i \in \{0, \ldots, p - 1\}$ be such that $a \in A_i$. Then $r(A(p)) = 1 - r(A_p) = pr(A_i) \geq p/a$.
4. The Theorem was stated for multisets to accommodate both the finite and the infinite case. In the following applications, however, we speak of IIDCS's, so the multisets of moduli are actually sets.

APPLICATIONS.

1. It is easy to see that the sum of the reciprocals of the numbers $2^i3^j, 2^i5^k, 3^j5^k, 2^i3^j5^k$ $(i, j, k \in \mathcal{P}^+)$ is 1. Beebee [**1**] asked whether there exists a saturated IIDCS whose moduli are these numbers. To have convenient notation, we let D_m, for $m \in \mathcal{P}^+$, denote the set of positive integers that are multiples of m and all of whose prime divisors divide m. With this notation, Beebee's set of moduli is

$$A = D_6 \cup D_{10} \cup D_{15} \cup D_{30},$$

and $r(D_m) = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$ for $m = 6, 10, 15, 30$, respectively. Suppose A is the set of moduli of a (necessarily saturated) DCS. Using the Theorem for $p = 2$, we get a partition of A into A_0, A_1, A_2. Assume w.l.o.g. that $10 \in A_0$. By

(iii) $D_6 \cap A_0 = \varnothing$, so $D_6 \subseteq A_1$. But $r(D_6) = \frac{1}{2}$ while $r(A_1) < \frac{1}{2}$ by (ii). Thus the answer to Beebee's question is negative.

2. The following are two other candidates to be the set of moduli of a saturated IIDCS:

$$A^* = D_6 \cup D_{15} \cup D_{20} \cup D_{30} \cup D_{42} \cup D_{105} \cup D_{140},$$
$$A^{**} = D_6 \cup D_{20} \cup D_{28} \cup D_{30} \cup D_{42} \cup D_{52} \cup D_{140} \cup D_{210}.$$

Each of these candidates can be rejected using Remark 3 above (for example, take $p = 7$ and $a = 42$ for A^*, and take $p = 13$ and $a = 52$ for A^{**}).

NOTE. Through correspondence we learned that John Selfridge has shown that Beebee's question has a negative answer by considering the residues mod 2 and mod 3. The same conclusion has been reached by Jamie Simpson, using a density argument. John Beebee has shown that our two additional systems of moduli cannot form a saturated IIDCS by using Selfridge's argument, and Doron Zeilberger has reached the same conclusion by a density argument.

REFERENCES

1. J. Beebee, Examples of infinite, incongruent exact covers, *Amer. Math. Monthly*, 95 (1988) 121–123.
2. J. Dewar, On finite and infinite covering sets, in: Proc. Washington State Univ. Conf. on Number Theory (Washington State Univ., Pullman, Washington, 1971), pp. 201–206.
3. P. Erdös, Problems and results on combinatorial number theory III, in: Number Theory Day, Proc. Conf. Rockefeller Univ. 1976, Lecture Notes in Math., vol. 626, pp. 43–49, Springer, Berlin, 1977.
4. C. E. Krukenberg, Covering sets of the integers, Ph.D. Thesis, Univ. of Ill., Urbana-Champaign, IL, 1971.
5. R. J. Simpson, Disjoint covering systems of congruences, *Amer. Math. Monthly* 94 (1987) 865–868.
6. S. K. Stein, Unions of arithmetic sequences, *Math. Annalen*, 134 (1958) 289–294.
7. D. Zeilberger, On a conjecture of R. J. Simpson about exact covering congruences, *Amer. Math. Monthly* 96 (1989) 243.
8. S. Znám, A survey of covering systems of congruences, *Acta Mathematica Universitatis Comenianae*, 40–41 (1982) 59–79.

A Moment Convergence Theorem

LAJOS TAKÁCS
Department of Mathematics and Statistics, Case Western Reserve University, Cleveland, OH 44106

In the theory of probability we frequently encounter the problem of finding the limit distribution of a sequence of real random variables. If the random variables have finite moments and their limits exist, then the following theorem seems appropriate.

THEOREM 1. *Let* $\{F_n(x)\}$ *be a sequence of distribution functions for which the moments*

$$M_r(n) = \int_{-\infty}^{\infty} x^r \, dF_n(x) \tag{1}$$

(iii) $D_6 \cap A_0 = \varnothing$, so $D_6 \subseteq A_1$. But $r(D_6) = \frac{1}{2}$ while $r(A_1) < \frac{1}{2}$ by (ii). Thus the answer to Beebee's question is negative.

2. The following are two other candidates to be the set of moduli of a saturated IIDCS:

$$A^* = D_6 \cup D_{15} \cup D_{20} \cup D_{30} \cup D_{42} \cup D_{105} \cup D_{140},$$
$$A^{**} = D_6 \cup D_{20} \cup D_{28} \cup D_{30} \cup D_{42} \cup D_{52} \cup D_{140} \cup D_{210}.$$

Each of these candidates can be rejected using Remark 3 above (for example, take $p = 7$ and $a = 42$ for A^*, and take $p = 13$ and $a = 52$ for A^{**}).

NOTE. Through correspondence we learned that John Selfridge has shown that Beebee's question has a negative answer by considering the residues mod 2 and mod 3. The same conclusion has been reached by Jamie Simpson, using a density argument. John Beebee has shown that our two additional systems of moduli cannot form a saturated IIDCS by using Selfridge's argument, and Doron Zeilberger has reached the same conclusion by a density argument.

REFERENCES

1. J. Beebee, Examples of infinite, incongruent exact covers, *Amer. Math. Monthly*, 95 (1988) 121–123.
2. J. Dewar, On finite and infinite covering sets, in: Proc. Washington State Univ. Conf. on Number Theory (Washington State Univ., Pullman, Washington, 1971), pp. 201–206.
3. P. Erdös, Problems and results on combinatorial number theory III, in: Number Theory Day, Proc. Conf. Rockefeller Univ. 1976, Lecture Notes in Math., vol. 626, pp. 43–49, Springer, Berlin, 1977.
4. C. E. Krukenberg, Covering sets of the integers, Ph.D. Thesis, Univ. of Ill., Urbana-Champaign, IL, 1971.
5. R. J. Simpson, Disjoint covering systems of congruences, *Amer. Math. Monthly* 94 (1987) 865–868.
6. S. K. Stein, Unions of arithmetic sequences, *Math. Annalen*, 134 (1958) 289–294.
7. D. Zeilberger, On a conjecture of R. J. Simpson about exact covering congruences, *Amer. Math. Monthly* 96 (1989) 243.
8. S. Znám, A survey of covering systems of congruences, *Acta Mathematica Universitatis Comenianae*, 40–41 (1982) 59–79.

A Moment Convergence Theorem

LAJOS TAKÁCS
Department of Mathematics and Statistics, Case Western Reserve University, Cleveland, OH 44106

In the theory of probability we frequently encounter the problem of finding the limit distribution of a sequence of real random variables. If the random variables have finite moments and their limits exist, then the following theorem seems appropriate.

THEOREM 1. *Let* $\{F_n(x)\}$ *be a sequence of distribution functions for which the moments*

$$M_r(n) = \int_{-\infty}^{\infty} x^r \, dF_n(x) \tag{1}$$

exist for all $r = 0, 1, 2, \ldots$. *Furthermore, let* $F(x)$ *be a distribution function for which the moments*

$$M_r = \int_{-\infty}^{\infty} x^r \, dF(x) \tag{2}$$

exist for all $r = 0, 1, 2, \ldots$. *If*

$$\lim_{n \to \infty} M_r(n) = M_r \tag{3}$$

for all $r = 0, 1, 2, \ldots$, *and if* $F(x)$ *is uniquely determined by the sequence of moments* $M_0, M_1, M_2, \ldots$, *then*

$$\lim_{n \to \infty} F_n(x) = F(x) \tag{4}$$

in every continuity point of $F(x)$.

In applying this theorem, we should determine the limits (3), prove that $\{M_r\}$ is a sequence of moments of a distribution function $F(x)$, and prove that $F(x)$ is uniquely determined by (2). In 1920 H. Hamburger [**7**] gave a necessary and sufficient condition for a sequence $\{M_r\}$ to be a moment sequence, and we have various sufficient conditions for the uniqueness of the moment problem. By a result of G. Pólya [**16**], $F(x)$ is uniquely determined by (2) if

$$\limsup_{m \to \infty} \frac{M_{2m}^{1/2m}}{m} < \infty, \tag{5}$$

and by a result of T. Carleman [**2**], $F(x)$ is uniquely determined by (2) if

$$\sum_{m=1}^{\infty} \frac{1}{M_{2m}^{1/2m}} = \infty. \tag{6}$$

See also J. A. Shohat and J. D. Tamarkin [**17**]. Even if $F(x)$ is uniquely determined by (2), we have to find ways to recover $F(x)$ from (2). There is no general method of finding $F(x)$. One possible way is to try to determine

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \, dF(x), \tag{7}$$

the Fourier-Stieltjes transform of $F(x)$, by $\{M_r\}$, and invert (7).

Theorem 1 has a long history. In the particular case when $F(x) = \Phi(x)$, the normal distribution function, in 1887 P. L. Chebyshev [**3**] proved that $\Phi(x)$ is uniquely determined by its sequence of moments and used the above theorem in his investigations in the theory of probability. However, he did not prove this theorem. For $F(x) = \Phi(x)$, the theorem was proved only in 1898 by A. A. Markov [**13**]. See also A. A. Markov [**14**]. In 1920 G. Pólya [**16**] proved Theorem 1 in the case when (5) is satisfied. In the stated general case Theorem 1 was proved in 1931 by M. Fréchet and J. Shohat [**6**]. Their proof is based on some general results of E. Helly [**9**].

In many applications we encounter a sequence of discrete random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ which take on only nonnegative integer values and we want to determine the limit distribution $\lim_{n \to \infty} \mathbf{P}\{\xi_n = k\}$, $k = 0, 1, 2, \ldots$, if it exists. In principle, we can apply Theorem 1, but we have to overcome all the difficulties mentioned earlier. On several occasions I used the method of binomial moments in

proving such limit theorems. See references [19], [20], [21], [22], [25], and [26]. This method has led to a general limit theorem whose proof is elementary and surprisingly simple. In addition it yields an explicit formula for the limit distribution. This note contains this theorem and its proof.

If ξ is a discrete random variable which takes on only nonnegative integer values, then its rth binomial moment is defined by

$$B_r = \mathbf{E}\left\{\binom{\xi}{r}\right\} = \sum_{k=r}^{\infty} \binom{k}{r}\mathbf{P}\{\xi = k\} \tag{8}$$

for $r = 0, 1, 2, \ldots$. The rth factorial moment of ξ is defined as $r!B_r$ for $r = 0, 1, 2, \ldots$. It seems binomial moments were introduced by G. F. Hardy [8 p. 59] for the first time in 1909 for the practical purpose of calculating moments simply by repeated additions. Factorial moments were used in 1921 by R. v. Mises [15] and in 1923 by J. F. Steffensen [18]. Steffensen proved that if $B_r = 0$ for $r > m$, then

$$\mathbf{P}\{\xi = k\} = \sum_{r=k}^{m} (-1)^{r-k}\binom{r}{k}B_r \tag{9}$$

for $k = 0, 1, \ldots, m$. The wide dissemination of the use of binomial moments is due to K. Jordan [10], [11].

THEOREM 2. *Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a sequence of discrete random variables taking on nonnegative integers only. Define $P_j(n) = \mathbf{P}\{\xi_n = j\}$. Let us suppose that the r-th binomial moment of ξ_n,*

$$B_r(n) = \mathbf{E}\left\{\binom{\xi_n}{r}\right\} = \sum_{j=r}^{\infty} \binom{j}{r}P_j(n), \tag{10}$$

is finite for all $r \geqq 0$ and $n \geqq 1$. If the limit

$$\lim_{n\to\infty} B_r(n) = B_r \tag{11}$$

exists for all $r \geqq j \geqq 0$, and if the limit

$$\lim_{n\to\infty} \sum_{r=j}^{n} (-1)^{r-j}\binom{r}{j}B_r = P_j \tag{12}$$

exists for $j \geqq 0$, then

$$\lim_{n\to\infty} P_j(n) = P_j. \tag{13}$$

Proof. Let us define a set of $m + 1$ polynomials of degree m in x by the following formulas

$$f_{m,j}(x) = \sum_{r=j}^{m} (-1)^{r-j}\binom{r}{j}\binom{x}{r} = \binom{x}{j}\sum_{r=j}^{m} (-1)^{r-j}\binom{x-j}{r-j}$$

$$= (-1)^{m-j}\binom{x}{j}\binom{x-j-1}{m-j} = \binom{x}{j}\binom{m-x}{m-j} \tag{14}$$

for $j = 0, 1, \ldots, m$ and $m \geqq 0$. The third equality can be proved simply by mathematical induction on m. We shall consider these polynomials only for

$x = 0, 1, 2, \ldots$. Let us define

$$\delta_j(x) = \begin{cases} 1 & \text{for } x = j, \\ 0 & \text{for } x \neq j. \end{cases} \tag{15}$$

Since $f_{m,j}(x) = \delta_j(x)$ if $0 \leqq x \leqq m$ and $0 \leqq j \leqq m$, and $(-1)^{m-j} f_{m,j}(x) \geqq 0$ if $x > j$ and $m \geqq j$, we have the inequalities

$$f_{j+2s+1,j}(x) \leqq \delta_j(x) \leqq f_{j+2s,j}(x) \tag{16}$$

whenever x, j, and s are nonnegative integers.

Putting $x = \xi_n$ in (16) and forming the expectation of each member, we obtain the following inequalities

$$\sum_{r=j}^{j+2s+1} (-1)^{r-j} \binom{r}{j} B_r(n) \leqq P_j(n) \leqq \sum_{r=j}^{j+2s} (-1)^{r-j} \binom{r}{j} B_r(n) \tag{17}$$

for $j \geqq 0$ and $s \geqq 0$. These inequalities are the generalizations of Bonferroni's inequalities. (See C. E. Bonferroni [**1**] or W. Feller [**5**, p. 110].) If we let $n \to \infty$ in (17), by (11) we obtain that

$$\sum_{r=j}^{j+2s+1} (-1)^{r-j} \binom{r}{j} B_r \leqq \liminf_{n\to\infty} P_j(n) \leqq \limsup_{n\to\infty} P_j(n) \leqq \sum_{r=j}^{j+2s} (-1)^{r-j} \binom{r}{j} B_r \tag{18}$$

holds for all $s \geqq 0$. If $s \to \infty$, then by (12), the extreme members in (18) tend to P_j and this proves that the limit (13) exists and is equal to P_j.

We note that if in Theorem 2, $B_r = a^r/r!$ for $r = 0, 1, 2, \ldots$ where $0 < a < \infty$, then the random variables ξ_n $(n = 1, 2, \ldots)$ have a limiting Poisson distribution, that is,

$$P_j = e^{-a} a^j / j! \tag{19}$$

for $j \geqq 0$. This result was proved by R. v. Mises [**15**] in 1921, and P. Erdös and A. Rényi [**4**, p. 27] in 1960 by using a more complicated approach than Theorem 2.

Finally, we note that if in Theorem 2, instead of (12), we assume only that the binomial moments $\{B_r\}$ uniquely determine the distribution $\{P_j\}$, then (13) remains valid, but (12) should be replaced by a more general formula.

According to a result of A. Lenard [**12**] the distribution $\{P_j\}$ is uniquely determined by the sequence $\{B_r\}$ if

$$\sum_{r=1}^{\infty} \frac{1}{B_r^{1/r}} = \infty. \tag{20}$$

If

$$\rho = \limsup_{r\to\infty} (B_r)^{1/r} < \infty, \tag{21}$$

then (20) is satisfied and $\{P_j\}$ is uniquely determined by the sequence of binomial moments B_r $(r = 0, 1, 2, \ldots)$; however, (12) should be replaced by the more general formula

$$P_j = \sum_{r=j}^{\infty} \binom{r}{j} (1+q)^{-(r+1)} \sum_{i=j}^{r} (-1)^{i-j} \binom{r-j}{i-j} q^{r-i} B_i \tag{22}$$

in which q is any nonnegative parameter satisfying the inequality $q > \rho^2 - 1$. In some particular cases we can choose $q > (\rho - 1)/2$. If $\rho < 1$, then the limit (12) exists and consequently (13) holds. In this case we can choose $q = 0$ and (22) reduces to (12). The above result is based on L. Takács [**23**], [**24**].

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

Why the Method of Undetermined Coefficients Works

CLAY C. ROSS, JR.
Department of Mathematics and Computer Science, University of the South, Sewanee, TN 37375

Introduction. As presented in many introductory differential equations books, the method of undetermined coefficients is somewhat mysterious. Zill [1] shows clearly how to use annihilators to solve undetermined coefficients problems, although he does not say why his procedure works. Hildebrand [2] hints at a similar method for difference equations. This note presents a simple theorem which explains why the method of undetermined coefficients works, both for differential equations and difference equations.

For our purposes, an nth order linear differential operator has the form $\mathbf{L}(y) = y^{(n)} + f_{n-1}(x)y^{(n-1)} + \cdots + f_1(x)y' + f_0(x)y$, and is defined on all $\mathbf{C}^n$ functions over some interval $\mathbf{I}$ on which the coefficients are smooth ($\mathbf{C}^\infty$), and an nth order linear difference operator has the form $\mathbf{M}(y_k) = y_{k+n} + a_{n-1}(k)y_{k+n-1} + \cdots + a_1(k)y_{k+1} + a_0(k)y_k$, where $a_0(k) \neq 0$, and y and every a_i is defined for all k on some interval $\mathbf{I}$ of nonnegative integers.

THEOREM. *If* $\mathbf{L}$ *and* $\mathbf{A}$ *are both linear differential or linear difference operators of positive order, and* f *is a function annihilated by* $\mathbf{A}$, *i.e., such that* $\mathbf{A}(f) = 0$, *then*

(a) *every solution of* $\mathbf{L}(y) = f$ *belongs to* $\mathrm{Ker}(\mathbf{AL})$;

(b) *in every subspace* $\mathbf{V}$ *of* $\mathrm{Ker}(\mathbf{AL})$ *complementary to* $\mathrm{Ker}(\mathbf{L})$ (*i.e., in every* $\mathbf{V}$ *such that* $\mathrm{Ker}(\mathbf{L}) \oplus \mathbf{V} = \mathrm{Ker}(\mathbf{AL})$), *there is a unique solution to* $\mathbf{L}(y) = f$.

Proof. Suppose for definiteness that $\mathbf{L}$ is of positive order n and $\mathbf{A}$ is of positive order m. Then $\mathbf{AL}$ is of order $n + m$. The parts of the theorem are proved as follows:

(a) Apply $\mathbf{A}$ to $\mathbf{L}(y) = f$: $\mathbf{AL}(y) = \mathbf{A}(f) = 0$.

(b) Let $\{y_1, y_2, \ldots, y_n\}$ be a basis for $\mathrm{Ker}(\mathbf{L})$. Since $\mathbf{AL}$ is of order $n + m$, extend this to a basis $\{y_1, y_2, \ldots, y_n, z_1, \ldots, z_m\}$ for $\mathrm{Ker}(\mathbf{AL})$, so that $\{z_1, \ldots, z_m\}$ is a basis for $\mathbf{V}$. Then $\mathbf{L}(z_1), \ldots, \mathbf{L}(z_m)$ all belong to $\mathrm{Ker}(\mathbf{A})$, since $z_1, \ldots, z_m$ all belong to $\mathrm{Ker}(\mathbf{AL})$. Moreover, $\mathbf{L}(z_1), \ldots, \mathbf{L}(z_m)$ are linearly independent, for if $\sum a_i \mathbf{L}(z_i) = 0$, then $\mathbf{L}(\sum a_i z_i) = 0$. This implies that $\sum a_i z_i$ is in $\mathrm{Ker}(\mathbf{L})$. But $\sum a_i z_i$ is also in $\mathbf{V}$, and hence is 0 since $\mathrm{Ker}(\mathbf{L})$ and $\mathbf{V}$ are complementary. Thus $a_i = 0$ for $i = 1, \ldots, m$.

Since $\mathrm{Ker}(\mathbf{A})$ has dimension m, $\{\mathbf{L}(z_1), \ldots, \mathbf{L}(z_m)\}$ is a basis for $\mathrm{Ker}(\mathbf{A})$. Thus f in $\mathrm{Ker}(\mathbf{A})$ implies f has the unique representation $f = \sum c_i \mathbf{L}(z_i)$. Thus $y = \sum c_i z_i$ in $\mathbf{V}$ is a solution to $\mathbf{L}(y) = f$, since $\mathbf{L}(y) = \sum c_i \mathbf{L}(z_i) = f$. The coefficients c_i are unique, so there is no other solution in $\mathbf{V}$. □

COROLLARY. *If* $\mathbf{L}$ *and* $\mathbf{A}$ *commute then* $\mathrm{Ker}(\mathbf{A})$ *is a subspace of* $\mathrm{Ker}(\mathbf{AL})$. *If, in addition,* $\mathrm{Ker}(\mathbf{L})$ *and* $\mathrm{Ker}(\mathbf{A})$ *have only* 0 *in common, then in* $\mathrm{Ker}(\mathbf{A})$ *there is a unique solution to* $\mathbf{L}(y) = f$.

Proof. Suppose **L** and **A** commute and z is in Ker(**A**) then $\mathbf{AL}(z) = \mathbf{LA}(z) = \mathbf{L}(0) = 0$, so Ker(**A**) is a subspace of Ker(**AL**). The rest follows from (b) of the theorem since now one can choose $\mathbf{V} = \text{Ker}(\mathbf{A})$. □

Observations. If **L** and **A** have constant coefficients, then **L** and **A** commute. This motivates the standard practice of undetermined coefficients; however, the complement of Ker(**L**) in Ker(**AL**) need have very little similarity to Ker(**A**), depending on just how large $\text{Ker}(\mathbf{L}) \cap \text{Ker}(\mathbf{A})$ is, and this is what often leaves students mystified.

It is not the fact that **L** and **A** are differential or difference operators which makes the theorem work, but the fact that for operators of these types, $\dim(\text{Ker}(\mathbf{AL})) = \dim(\text{Ker}(\mathbf{A})) + \dim(\text{Ker}(\mathbf{L}))$.

The theorem explains exactly which nonhomogeneous linear differential or difference equations permit finding a particular solution by the method of undetermined coefficients: the right-hand side must be annihilated by some linear differential or difference operator of positive order. In the constant coefficients case for either topic, it is clear from the theory how to obtain **A** and all of the basis functions specified by the theorem. In fact, the bases for Ker(**L**) and Ker(**AL**) so obtained will always give a basis for Ker(**L**) as a subset of a basis for Ker(**AL**). This means that one can immediately identify a basis for a subspace complementary to Ker(**L**) as those basis elements in Ker(**AL**) which were not in Ker(**L**). One then knows the form of a particular solution since the existence of a unique solution in this complementary subspace is guaranteed. These ideas are illustrated by the following examples.

Examples. Consider $\mathbf{L}(y) = (D - 2)(D - 4)(D + 3)^2 y = 5x^3 e^{-3x}$, where $D \equiv d/dx$. Then $\mathbf{A} = (D + 3)^4$ is an annihilator of e^{-3x}, xe^{-3x}, and $x^2 e^{-3x}$, as well as $x^3 e^{-3x}$, and the various bases described in the theorem are:

$$(\mathbf{AL}): \quad \{e^{2x}, e^{4x}, e^{-3x}, xe^{-3x}, x^2 e^{-3x}, x^3 e^{-3x}, x^4 e^{-3x}, x^5 e^{-3x}\}$$

$$(\mathbf{L}): \quad \{e^{2x}, e^{4x}, e^{-3x}, xe^{-3x}\}$$

$$(\mathbf{A}): \quad \{e^{-3x}, xe^{-3x}, x^2 e^{-3x}, x^3 e^{-3x}\}$$

$$(\mathbf{V}): \quad \{x^2 e^{-3x}, x^3 e^{-3x}, x^4 e^{-3x}, x^5 e^{-3x}\}$$

Now assume a particular solution of the form $y_p = c_1 x^2 e^{-3x} + c_2 x^3 e^{-3x} + c_3 x^4 e^{-3x} + c_4 x^5 e^{-3x}$. Note that the form of the assumed solution differs markedly from the form of the basis for Ker(**A**) because $\dim(\text{Ker}(\mathbf{L}) \cap \text{Ker}(\mathbf{A})) = 2$.

Analogously, let $\mathbb{L}(y_k) = (E - 2)(E - 4)(E + 3)^2 y_k = 5k^3(1/3)^k$, where $E^s y_k \equiv y_{k+s}$. Then $\mathbb{A} = (E - 1/3)^4$ is an annihilator of $(1/3)^k$, $k(1/3)^k$, and $k^2(1/3)^k$, as well as $k^3(1/3)^k$, and the various bases described in the theorem are:

$$(\mathbf{AL}): \quad \{2^k, 4^k, (1/3)^k, k(1/3)^k, k^2(1/3)^k, k^3(1/3)^k, k^4(1/3)^k, k^5(1/3)^k\}$$

$$(\mathbf{L}): \quad \{2^k, 4^k, (1/3)^k, k(1/3)^k\}$$

$$(\mathbf{A}): \quad \{(1/3)^k, k(1/3)^k, k^2(1/3)^k, k^3(1/3)^k\}$$

$$(\mathbf{V}): \quad \{k^2(1/3)^k, k^3(1/3)^k, k^4(1/3)^k, k^5(1/3)^k\}$$

Now assume a particular solution of the form $y_p = c_1k^2(1/3)^k + c_2k^3(1/3)^k + c_3k^4(1/3)^k + c_4k^5(1/3)^k$. Note again that the form of the assumed solutions differs from the form of the basis for Ker(**A**) because $\dim(\text{Ker}(\mathbf{L}) \cap \text{Ker}(\mathbf{A})) = 2$.

Acknowledgement. The author expresses his appreciation to Stephen E. Puckette, whose incisive understanding at a critical moment made everything come together.

REFERENCES

1. Dennis G. Zill, A First Course in Differential Equations with Applications, 4th edition, Prindle, Weber & Schmidt, Boston, 1989, p. 172–177.
2. Francis B. Hildebrand, Finite-Difference Equations and Simulations, Prentice-Hall, Englewood Cliffs, New Jersey, 1968, pp. 32–33, 90.

A Freshman Calculus Proof that Real Quartics Factor

JOHN ISBELL
Department of Mathematics, State University of New York, Buffalo, NY 14214

This note gives a proof that is technically within the grasp of second semester freshmen, and as natural (I think) as the usual proofs that $\lim_{x \to 0} \sin x/x = 1$, that a fourth degree real polynomial factors into two real quadratic factors. (So, by the quadratic formula, it has four complex roots.) The required calculus is the intermediate value theorem, in the form: If $f(x)$ is a polynomial with positive leading coefficient and $f(0) < 0$, then there is $r > 0$ such that $f(r) = 0$.

Two further selling points. (1) The idea is rather like completing the square. Just as every real quadratic $x^2 + ax + b$ is $l(x)^2 \pm c^2$, where $l(x)$ is linear and c is constant, so a quartic $x^4 + ax^3 + bx^2 + cx + d$ is $q(x)^2 - l(x)^2$, where $q(x)$ is quadratic and $l(x)$ is linear. (2) The proof turns on associated cubics and quadratics. To get the formula for the solution of the quartic would be much harder, because this proof branches into three cases; but it certainly shows that any particular quartic equation has roots expressible in terms of roots of explicitly given cubics and quadratics (only one cubic).

1. Derivation. Consider a real quartic polynomial $Q(x)$ which we write as $x^4 + 2Ax^3 + Bx^2 + Cx + D$. We shall show that $Q(x) = g(x)^2 - h(x)^2$, where $g(x) = x^2 + Ax + t$ and $h(x) = rx + s$, for some real numbers r, s, t. This will give the factorization $Q(x) = (g(x) + h(x))(g(x) - h(x))$. Equating coefficients in $g(x)^2 - h(x)^2 = Q(x)$ gives

$$\begin{aligned} A^2 + 2t - r^2 &= B \\ 2At - 2rs &= C \\ t^2 - s^2 &= D. \end{aligned} \tag{1}$$

Solving the first of these for t,

$$t = \tfrac{1}{2}(r^2 + B - A^2). \tag{T}$$

This is equivalent to the first equation of (1), so we may as well substitute in the

from a movie showing the complex roots of $Q_t(x)$ for $t = 1, 2, 3, 4$. $Q_1(x)$ has roots $\pm i, 1, 2$. The imaginaries move to 0, giving $Q_2(x)$ roots $0, 0, 1, 2$. Then the roots stay real and three of them stay constant while one grows, giving $Q_3(x)$ roots $0, 1, 1, 2$. Finally the double root separates to $1 + i$ and $1 - i$, which with 0 and 2 are the roots of $Q_4(x)$. $Q_1(x)$ and $Q_4(x)$ are uniquely factorable into two real quadratics, and the factors of $Q_1(x)$ can't be changed continuously into those of $Q_4(x)$ by way of real quadratic factors of $Q_t(x)$, for t in $(1, 4)$.

On the Right-Hand Derivative of a Certain Integral Function

JOHN KLIPPERT
Department of Mathematics and Computer Science, James Madison University, Harrisonburg, VA 22807

Suppose f is a real-valued function defined on the interval $[0, 1]$ such that

a) f is bounded on $[0, 1]$,

b) f is continuous on $(0, 1]$,

and

c) $\lim_{x \to 0^+} f(x)$ does not exist.

Since f is bounded and has a single discontinuity on $[0, 1]$, it is Riemann integrable there. Let $F(x) = \int_0^x f(t)\,dt$, $x \in [0, 1]$. Since F is continuous on $[0, 1]$ and differentiable on $(0, 1]$ (see [**2**, p. 147]), a natural (and interesting) question for the student of "advanced calculus" to investigate is:

Does

$$F'_+(0) = \lim_{x \to 0^+} \frac{F(x) - F(0)}{x - 0} = \lim_{x \to 0^+} \frac{\int_0^x f(t)\,dt}{x}$$

also exist?

At first, the student is tempted to misapply l'Hospital's Rule and conclude that the limit does not exist. To show that this need not be the case, we can take $f(x) = \sin 1/x$ for $x \in (0, 1]$ with $f(0) = 0$ (any other value for f at 0 will do just as well), and appeal to the elegant argument given by Steve Ricci in [**1**] that

$$\lim_{x \to 0^+} \frac{\int_0^x f(t)\,dt}{x} = 0.$$

The argument, which relies upon integration by parts, is well within the grasp of the student. Based upon this example, the student may then wonder whether or not the limit in question *always* exists. Providing an example such that (a)–(c) hold, but $\lim_{x \to 0^+} \int_0^x f(t)\,dt/x$ does not exist affords the student a nice challenge. Such an example (which again appeals to integration by parts) is offered here.

Let us put $g(x) = \sin(\ln x)$ for $x \in (0, 1]$ with $g(0) = 0$. Note that (a)–(c) hold for g. Using integration by parts, we find that the function

$$F(x) = \begin{cases} \dfrac{x}{2}(\sin(\ln x) - \cos(\ln x)), & x \in (0, 1] \\ 0, & x = 0 \end{cases}$$

from a movie showing the complex roots of $Q_t(x)$ for $t = 1, 2, 3, 4$. $Q_1(x)$ has roots $\pm i, 1, 2$. The imaginaries move to 0, giving $Q_2(x)$ roots $0, 0, 1, 2$. Then the roots stay real and three of them stay constant while one grows, giving $Q_3(x)$ roots $0, 1, 1, 2$. Finally the double root separates to $1 + i$ and $1 - i$, which with 0 and 2 are the roots of $Q_4(x)$. $Q_1(x)$ and $Q_4(x)$ are uniquely factorable into two real quadratics, and the factors of $Q_1(x)$ can't be changed continuously into those of $Q_4(x)$ by way of real quadratic factors of $Q_t(x)$, for t in $(1, 4)$.

On the Right-Hand Derivative of a Certain Integral Function

JOHN KLIPPERT
Department of Mathematics and Computer Science, James Madison University, Harrisonburg, VA 22807

Suppose f is a real-valued function defined on the interval $[0, 1]$ such that

a) f is bounded on $[0, 1]$,

b) f is continuous on $(0, 1]$,

and

c) $\lim_{x \to 0^+} f(x)$ does not exist.

Since f is bounded and has a single discontinuity on $[0, 1]$, it is Riemann integrable there. Let $F(x) = \int_0^x f(t)\,dt$, $x \in [0, 1]$. Since F is continuous on $[0, 1]$ and differentiable on $(0, 1]$ (see [**2**, p. 147]), a natural (and interesting) question for the student of "advanced calculus" to investigate is:

Does

$$F'_+(0) = \lim_{x \to 0^+} \frac{F(x) - F(0)}{x - 0} = \lim_{x \to 0^+} \frac{\int_0^x f(t)\,dt}{x}$$

also exist?

At first, the student is tempted to misapply l'Hospital's Rule and conclude that the limit does not exist. To show that this need not be the case, we can take $f(x) = \sin 1/x$ for $x \in (0, 1]$ with $f(0) = 0$ (any other value for f at 0 will do just as well), and appeal to the elegant argument given by Steve Ricci in [**1**] that

$$\lim_{x \to 0^+} \frac{\int_0^x f(t)\,dt}{x} = 0.$$

The argument, which relies upon integration by parts, is well within the grasp of the student. Based upon this example, the student may then wonder whether or not the limit in question *always* exists. Providing an example such that (a)–(c) hold, but $\lim_{x \to 0^+} \int_0^x f(t)\,dt/x$ does not exist affords the student a nice challenge. Such an example (which again appeals to integration by parts) is offered here.

Let us put $g(x) = \sin(\ln x)$ for $x \in (0, 1]$ with $g(0) = 0$. Note that (a)–(c) hold for g. Using integration by parts, we find that the function

$$F(x) = \begin{cases} \dfrac{x}{2}(\sin(\ln x) - \cos(\ln x)), & x \in (0, 1] \\ 0, & x = 0 \end{cases}$$

is an antiderivative of g on $(0, 1]$. Thus, since F is continuous on $[0, 1]$ (and g is integrable of $[0, 1]$), we have by the Fundamental Theorem of Calculus (for $x \in (0, 1)$] that $\int_0^x g(t)\,dt = F(x) - F(0) = F(x)$ and the equation holds trivially for $x = 0$. Putting $x_n = e^{-n\pi}$, $n \in \mathbb{N}$, we find that although $\lim_{n\to\infty} x_n = 0$,

$$\lim_{n\to\infty} \frac{\int_0^{x_n} g(t)\,dt}{x_n} = \lim_{n\to\infty} \frac{\sin(-n\pi) - \cos(-n\pi)}{2} = \lim_{n\to\infty} -\frac{\cos(n\pi)}{2} = \lim_{n\to\infty} \frac{(-1)^{n+1}}{2}$$

which does not exist and, hence, $\lim_{x\to 0^+} \int_0^x g(t)\,dt/x$ does not exist.

In comparing f with g, notice that the intervals between consecutive zeros of g are exponentially smaller than those for f. For f, the length of the interval

$$\left[\frac{1}{(n+1)\pi}, \frac{1}{n\pi}\right]$$

is $l_n(f) = 1/n(n+1)\pi$, while for g, the length of the interval $[1/e^{(n+1)\pi}, 1/e^{n\pi}]$ is $l_n(g) = (e^{\pi} - 1)/e^{(n+1)\pi}$ and we observe that $\lim_{n\to\infty} l_n(g)/l_n(f) = 0$. Thus the more highly oscillatory nature of g appears to prevent the right differentiability of its integral function at the origin.

Finally, in light of these (and other) examples, the conjecture can be made (although the author is by no means certain of its validity) that if in addition to conditions (a)–(c), the zeros of f form a null sequence of distinct terms whose associated sequence of distances between consecutive zeros is asymptotically proportional to $1/n^{\alpha}$ for some $\alpha > 1$, then $F'_+(0) = 0$.

REFERENCES

1. S. Ricci, Solution to proposed problem 1112, *Mathematics Magazine*, 55 (1982) 48.
2. W. F. Trench, Advanced Calculus, Harper and Row, New York, 1978.

Geographical Boundary Extrema

HelenJane Armstrong
Department of Geography, University of Florida, Gainesville, FL 32611

Roger Herz-Fischler
Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6

H. Thurston [**1**] has given Florida as an example of a geographical entity that has its highest elevation on the boundary. However, upon checking a detailed map [**2**] one finds, contrary to the claim, that the highest point in Florida is not on the border with Alabama, but rather is located approximately half a mile south of the border (just south of the village of Lakewood, Walton County; T. 6N, R. 20W, section 30). In [**3**, p. 4] the maximum elevation is given as 345 feet whereas the values at the border appear from the map to be bounded above by 340 feet.

is an antiderivative of g on $(0, 1]$. Thus, since F is continuous on $[0, 1]$ (and g is integrable of $[0, 1]$), we have by the Fundamental Theorem of Calculus (for $x \in (0, 1)$] that $\int_0^x g(t)\,dt = F(x) - F(0) = F(x)$ and the equation holds trivially for $x = 0$. Putting $x_n = e^{-n\pi}$, $n \in \mathbb{N}$, we find that although $\lim_{n\to\infty} x_n = 0$,

$$\lim_{n\to\infty} \frac{\int_0^{x_n} g(t)\,dt}{x_n} = \lim_{n\to\infty} \frac{\sin(-n\pi) - \cos(-n\pi)}{2} = \lim_{n\to\infty} -\frac{\cos(n\pi)}{2} = \lim_{n\to\infty} \frac{(-1)^{n+1}}{2}$$

which does not exist and, hence, $\lim_{x\to 0^+} \int_0^x g(t)\,dt/x$ does not exist.

In comparing f with g, notice that the intervals between consecutive zeros of g are exponentially smaller than those for f. For f, the length of the interval

$$\left[\frac{1}{(n+1)\pi}, \frac{1}{n\pi}\right]$$

is $l_n(f) = 1/n(n+1)\pi$, while for g, the length of the interval $[1/e^{(n+1)\pi}, 1/e^{n\pi}]$ is $l_n(g) = (e^{\pi} - 1)/e^{(n+1)\pi}$ and we observe that $\lim_{n\to\infty} l_n(g)/l_n(f) = 0$. Thus the more highly oscillatory nature of g appears to prevent the right differentiability of its integral function at the origin.

Finally, in light of these (and other) examples, the conjecture can be made (although the author is by no means certain of its validity) that if in addition to conditions (a)–(c), the zeros of f form a null sequence of distinct terms whose associated sequence of distances between consecutive zeros is asymptotically proportional to $1/n^{\alpha}$ for some $\alpha > 1$, then $F'_+(0) = 0$.

REFERENCES

1. S. Ricci, Solution to proposed problem 1112, *Mathematics Magazine*, 55 (1982) 48.
2. W. F. Trench, Advanced Calculus, Harper and Row, New York, 1978.

Geographical Boundary Extrema

HelenJane Armstrong
Department of Geography, University of Florida, Gainesville, FL 32611

Roger Herz-Fischler
Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6

H. Thurston [**1**] has given Florida as an example of a geographical entity that has its highest elevation on the boundary. However, upon checking a detailed map [**2**] one finds, contrary to the claim, that the highest point in Florida is not on the border with Alabama, but rather is located approximately half a mile south of the border (just south of the village of Lakewood, Walton County; T. 6N, R. 20W, section 30). In [**3**, p. 4] the maximum elevation is given as 345 feet whereas the values at the border appear from the map to be bounded above by 340 feet.

For a valid example consider the province of Alberta. Since the B.C.–Alberta border in the vicinity of 12,294-foot Mt. Columbia is determined by the watershed line [4], and since Alberta has no interior points, or other boundary points, that high, Mt. Columbia's summit is a boundary maximum for Alberta.

British Columbia itself presents an interesting example that emphasizes the importance of the domain of definition in extremal problems. Following the decisions of the 1903 Alaska Boundary Tribunal the International Boundary Commission published a report [5] dealing with the survey and demarcation of the boundary between B.C. and Alaska. Since the placing of a permanent marker on the peak of Mt. Fairweather (approximate height 15,230 feet) was not feasible, the report [5, p. 175, boundary point 164] states only: "Station mark: the highest part of the snow cap as it was in 1907." This statement provides the present legal description of the fixed boundary point. Coordinates of this snow cap had been determined from surveys, but since even for the key boundary point at Mt. St. Elias the calculations of two surveying parties differ by about 200 feet [5, p. 134], the actual location of the boundary point at Mt. Fairweather must be considered, in surveying terms, to be only very approximately known.

Because snow accumulation since 1907 has surely changed the location of the highest part of the snow cap, the station mark no longer exists in a physical sense. Further, because of the inaccurate survey, we cannot state from a practical viewpoint—unless the drift has been very large—whether the present summit of Mt. Fairweather is legally in B.C. or in Alaska. In the former case we would have an interior absolute maximum for B.C.; in the latter case the present legal absolute maximum for B.C. is somewhere on Fairweather's slopes. It should be noted that because of the watershed definition, this situation cannot occur on the B.C.–Alberta border and indeed the border has been recently redrawn after it was noticed that a body of water was flowing the "wrong way!"

Acknowledgements. The authors wish to thank the staff at Canada's Ministry of Energy, Mines and Resources, and the referee.

REFERENCES

1. H. Thurston, So-called 'exceptional' extremum problems, this MONTHLY, 68, (1961) 650–652. Reprinted in Selected Papers on Calculus, The Mathematical Association of America, 1969, 268–270.
2. U.S. Geological Survey, 7.5 minute series (topographic 1 : 24,000) Paxton, Florida–Alabama, revised 1987.
3. U.S. Geological Survey, Elevations and Distances in the United States, U.S. Government printing Office, 1983.
4. Office of the Surveyor General, Report of the Commission Appointed to Delimit the Boundary Between the Provinces of Alberta and British Columbia, 4 parts, Ottawa, 1917, 1924, 1925, 1955.
5. International Boundary commission [Canada, U.S.]. Joint Report Upon the Survey and Demarcation of the Boundary between Canada and the United States from Tongass Passage to Mount St. Elias., Ottawa, Department of Mines and Technical Surveys, 1952.

E 3460. *Proposed by E. Ehrhart, University of Strasbourg, France.*

(a) Suppose we have n mutually perpendicular chords through a point P interior to a sphere S in n-dimensional Euclidean space. Prove that the sum of the squares of the lengths of these chords depends only on the radius r of the sphere and the distance d from P to the center of the sphere.

(b) More generally suppose $1 \leqslant k \leqslant n$. Each set of k of the n mutually perpendicular chords through P given in (a) determines a k-dimensional affine subspace. Prove that the sum of the $(2/k)$-th powers of the k-dimensional measures of the cross-sections of S made by these $\binom{n}{k}$ affine subspaces depends only on r and d.

E 3461. *Proposed by David Callan, University of Wisconsin, Madison, WI.*

Suppose r is a rational number but not an integer. It is known that $\tan(r\pi/2)$ is an algebraic number. (Cf. Ivan Niven, *Irrational Numbers*, Carus Mathematical Monographs No. 11, pp. 37–41.) Find the smallest positive integer k_r such that $k_r \tan(r\pi/2)$ is an algebraic integer.

E 3462. *Proposed by J. J. Rotman, University of Illinois at Urbana-Champaign.*

Prove that any connected simple graph with an even number of edges has an orientation (assignment of direction to each edge) such that the number of edges leaving each vertex is even.

SOLUTIONS OF ELEMENTARY PROBLEMS

Extremal Morley Triangles

E 3307 [1989, 154]. *Proposed by Peter Andrews, Wilfrid Laurier University, Waterloo, Ontario, Murray Klamkin, University of Alberta, Edmonton, Alberta, and Edward T. H. Wang, Wilfrid Laurier University.*

The celebrated Morley triangle of a given triangle ABC is the equilateral triangle whose vertices are the intersections of adjacent pairs of internal angle trisectors of ABC. If s, R, r, F and s_M, R_M, r_M, F_M are the semiperimeter, the circumradius, the inradius, and the area, respectively, of ABC and its Morley triangle, determine the maximum of (i) s_M/s, (ii) R_M/R, (iii) r_M/r, and (iv) F_M/F.

Editorial remark. The first, second, and fourth ratios considered in the problem achieve their maxima when $A = B = C = \pi/3$ and have greatest lower bound zero. The third ratio, r_M/r, achieves a positive minimum when $A = B = C = \pi/3$ and has least upper bound equal to $2/9$. The calculations of these extrema are given in their approximate order of difficulty.

Solution of (ii) by the proposers. We prove that

$$R_M/R \leqslant (8/\sqrt{3})\sin^3(\pi/9),$$

with equality only when the original triangle is equilateral. We require the known fact that the Morley triangle has side length $8R\sin(A/3)\sin(B/3)\sin(C/3)$. (Cf.

[1], [2], or [3] in the list of references.) Since the circumradius of an equilateral triangle is $1/\sqrt{3}$ times its side-length, we have

$$R_M/R = (8/\sqrt{3})\sin(A/3)\sin(B/3)\sin(C/3).$$

Now if $g_2(x) = \log\sin(x/3)$, then $g_2''(x) = -\{3\sin(x/3)\}^{-2}$. Since $g_2''(x) < 0$ on $(0, \pi)$, we have by concavity

$$g_2(A) + g_2(B) + g_2(C) \leqslant 3g_2(\{A + B + C\}/3) = 3g_2(\pi/3),$$

or

$$\sin(A/3)\sin(B/3)\sin(C/3) \leqslant \sin^3(\pi/9),$$

with equality only if $A = B = C = \pi/3$. Thus the claimed result follows.

Solution of (iii) by the proposers. We prove that

$$(8/\sqrt{3})\sin^3(\pi/9) \leqslant r_M/r < 2/9,$$

with equality on the left only if the original triangle is equilateral. We begin by using the formulas $r_M = R_M/2$ and $r = 4R\sin(A/2)\sin(B/2)\sin(C/2)$ to obtain

$$r_M/r = (1/8)(R_M/R)\big/\{\sin(A/2)\sin(B/2)\sin(C/2)\}$$

or, by the solution of part (ii),

$$\frac{r_M}{r} = \frac{1}{\sqrt{3}}\,\frac{\sin(A/3)\sin(B/3)\sin(C/3)}{\sin(A/2)\sin(B/2)\sin(C/2)}.$$

Now if $g_3(x) = \log\{\sin(x/3)/\sin(x/2)\}$, then

$$\begin{aligned} g_3''(x) &= -\{3\sin(x/3)\}^{-2} + \{2\sin(x/2)\}^{-2} \\ &= \frac{\{\sin(x/3)/(x/3)\}^2 - \{\sin(x/2)/(x/2)\}^2}{\{6x^{-1}\sin(x/3)\sin(x/2)\}^2}. \end{aligned}$$

Since $(\sin t)/t$ is strictly decreasing on $(0, \pi/2)$, it follows that $g_3''(x) > 0$ for $0 < x < \pi$. Hence by convexity

$$g_3(A) + g_3(B) + g_3(C) \geqslant 3g_3(\{A + B + C\}/3) = 3g_3(\pi/3)$$

or

$$\frac{\sin(A/3)\sin(B/3)\sin(C/3)}{\sin(A/2)\sin(B/2)\sin(C/2)} \geqslant \left\{\frac{\sin(\pi/9)}{\sin(\pi/6)}\right\}^3 = 8\sin^3(\pi/9).$$

Thus we have the lower bound $r_M/r \geqslant (8/\sqrt{3})\sin^3(\pi/9)$, with equality only if $A = B = C = \pi/3$.

To get an upper bound we note that if $0 < A \leqslant C \leqslant B < \pi$ and if ε is a small positive number, then

$$\begin{aligned} &g_3(A - \varepsilon) + g_3(B + \varepsilon) - g_3(A) - g_3(B) \\ &\quad = \int_0^{\varepsilon}\{g_3'(B + t) - g_3'(A - t)\}\,dt > 0, \end{aligned}$$

since g_3' is strictly increasing on $(0, \pi)$. Thus r_M/r increases if we increase the largest angle of the triangle and decrease another angle by the same amount.

Since $g_1'(x) > 0$ for $0 < x < \pi$, we have

$$\frac{\sin(A/3)}{\cos(A/2)} < \frac{\sin((\pi - \theta_1)/3)}{\cos((\pi - \theta_1)/2)} < .76,$$

$$\frac{\sin(B/3)}{\cos(B/2)} < \frac{\sin(\pi/6)}{\cos(\pi/4)} = \sqrt{2}/2,$$

$$\frac{\sin(C/3)}{\cos(C/2)} < \frac{\sin((\pi - 2\theta_1)/3)}{\cos((\pi - 2\theta_1)/2)} < .037.$$

It follows that in case (b) we have $s_M/s < .06$, a bound smaller than that found in case (a).

Finally suppose that (c) holds. Say $A > \theta_1$ and $B, C < \theta_1$. Since $g_1''(x) < 0$ on $(0, \theta_1)$, we have by concavity

$$g_1(B) + g_1(C) \leqslant 2g_1((B + C)/2)$$

with strict inequality for $B \neq C$. Thus we may assume $B = C$. Since $A > \theta_1$,

$$B = \frac{1}{2}(\pi - A) < \frac{1}{2}(\pi - \theta_1) = .81275\ldots .$$

If we set $f_1(x) = 2g_1(x) + g_1(\pi - 2x)$, we have $s_M/s = 3\exp f_1(B)$. Now $f_1'(x) = 2g_1'(x) - 2g_1'(\pi - 2x)$ and we claim that $f_1'(x) > 0$ for $0 < x \leqslant (\pi - \theta_1)/2$. Indeed,

$$g_1'(x) = \frac{1}{2}\tan\frac{x}{2} + \frac{1}{3}\cot\frac{x}{3} > \frac{x}{4} + \frac{1 - x^2/18}{3(x/3)} > \frac{1}{x}$$

and

$$g_1'(\pi - 2x) = \frac{1}{2}\cot x + \frac{1}{3}\tan\left(\frac{\pi}{6} + \frac{2x}{3}\right) < \frac{1}{2x} + \frac{1}{3}\tan\left(\frac{\pi}{6} + \frac{2x}{3}\right).$$

For $0 < x \leqslant .82$ we have

$$f_1'(x) > \frac{1}{x} - \frac{2}{3}\tan\left(\frac{\pi}{6} + \frac{2x}{3}\right) > \frac{1}{.82} - \frac{2}{3}\tan\left(\frac{\pi}{6} + \frac{1.64}{3}\right) > .0007.$$

Thus, in this case

$$s_M/s < 3\exp f\left(\frac{\pi}{2} - \frac{\theta_1}{2}\right) = .16976\ldots < .18479\ldots ,$$

and hence the maximum of s_M/s can occur only in case (a) with $A = B = C$. Thus the claimed result $s_M/s \leqslant (8/\sqrt{3})\sin^3(\pi/9)$ is established.

Solution of (iv) by the editors. We prove that

$$F_M/F \leqslant (64/3)\sin^6(\pi/9) = 0.034148\ldots ,$$

with equality iff ABC is equilateral. Our argument is similar to that for (i), but is a little more complicated in case (c).

We have

$$F = abc/(4R) = 2R^2 \sin A \sin B \sin C$$

and

$$F_M = (\sqrt{3}/4)\{8R\sin(A/3)\sin(B/3)\sin(C/3)\}^2,$$

so that

$$\frac{F_M}{F} = 8\sqrt{3}\,\frac{\{\sin(A/3)\sin(B/3)\sin(C/3)\}^2}{\sin A \sin B \sin C}.$$

If $g_4(x) = \log\{\sin^2(x/3)/\sin x\}$, then $g_4'(x) = (2/3)\cot(x/3) - \cot x$ and

$$g_4''(x) = \{\sin x\}^{-2} - 2\{3\sin(x/3)\}^{-2}.$$

Now $g_4''(x)$ is negative for

$$0 < x < 3\arcsin\left(\left\{3(2-\sqrt{2})/8\right\}^{1/2}\right) = 1.46342\ldots = \theta_4$$

and is positive for $\theta_4 < x < \pi$. We treat three cases.

(a) If each of $A, B, C \leqslant \theta_4$, then, since $g_4''(x) < 0$ for $0 < x < \theta_4$, we have by concavity

$$g_4(A) + g_4(B) + g_4(C) \leqslant 3g_4((A+B+C)/3) = 3g_4(\pi/3)$$

with equality iff $A = B = C$. Thus in this case

$$F_M/F \leqslant 8\sqrt{3}\,\exp\{3g_4(\pi/3)\} = (64/3)\sin^6(\pi/9).$$

(b) If $A \geqslant B > \theta_4$, then $A < \pi - \theta_4$, $B < \pi/2$, and $C < \pi - 2\theta_4$. Now $g_4'(x) > 0$ for $0 < x < \pi/2$, since $\tan(x/3) < (1/3)\tan x$ there, and $g_4'(x) > 0$ for $\pi/2 \leqslant x < \pi$, since $\cot(x/3) > 0$ and $-\cot x \geqslant 0$ there. It follows that in this case

$$F_M/F \leqslant 8\sqrt{3}\,\frac{\{\sin((\pi-\theta_4)/3)\sin(\pi/6)\sin((\pi-2\theta_4)/3)\}^2}{\sin(\pi-\theta_4)\sin(\pi/2)\sin(\pi-2\theta_4)}$$

$$= .023552\ldots.$$

(c) If $A > \theta_4 \geqslant B, C$, then, since $g_4''(x) < 0$ for $0 < x < \theta_4$, we have by concavity

$$g_4(B) + g_4(C) \leqslant 2g_4((B+C)/2)$$

with equality iff $B = C$. Thus we may assume that $B = C$ and $B = (\pi - A)/2 < (\pi - \theta_4)/2 = .83908\ldots$. Let $f_4(x) = 2g_4(x) + g_4(\pi - 2x)$. We shall show that

$$g_4'(x) > g_4'(\pi - 2x), \quad 0 < x \leqslant (\pi - \theta_4)/2, \qquad (*)$$

and conclude that f_4 is increasing on this range.

Suppose first that $x \in I = (.7, (\pi - \theta 4)/2]$. Since g_4' is decreasing for $0 < x < \theta_4$ and increasing for $\theta_4 < x < \pi$, we have

$$g_4'(x) \geqslant g_4'((\pi-\theta_4)/2) = 1.42306\ldots$$
$$> 1.18884\ldots = g_4'(\pi - 1.4) \geqslant g'(\pi - 2x),$$

so $(*)$ holds for $x \in I$.

Next, suppose that $x \in J = (0, .7]$. We begin by showing that

$$3\sqrt{2}\,\sin(x/3) < \sin(2x), \quad x \in J.$$

Indeed

$$\frac{\sin(2x)}{\sin(x/3)} = 2\cos\left(\frac{5x}{3}\right) + 2\cos x + 2\cos\left(\frac{x}{3}\right)$$

is clearly decreasing on J and we have

$$\sin(2x)/\sin(x/3) > \sin 1.4/\sin(.7/3) = 4.26192\ldots > 3\sqrt{2}\,.$$

Now $g_4'(x) > (1/3)\cot(x/3)$, since $\tan(x/3) < (\tan x)/3$ for $x \in J$, and also we have

$$g_4'(\pi - 2x) = \frac{2}{3}\tan\left(\frac{\pi}{6} + \frac{2x}{3}\right) + \cot(2x)\,.$$

Let $h(x) = (1/3)\cot(x/3) - \cot(2x)$. We prove $(*)$ for $x \in J$ by showing that

$$h(x) > (2/3)\tan(\pi/6 + 2x/3) \qquad (**)$$

holds on this range. We have seen above that

$$h'(x) = 2\csc^2(2x) - \frac{1}{9}\csc^2\left(\frac{x}{3}\right) = \left(\frac{\sqrt{2}}{\sin 2x}\right)^2 - \left(\frac{1}{3\sin(x/3)}\right)^2 < 0$$

for $x \in J$. Thus $h(x) \geqslant h(.7) = 1.23007\ldots$. On the other hand, if $x \in J$, we have

$$(2/3)\tan(\pi/6 + 2x/3) \leqslant (2/3)\tan(\pi/6 + 1.4/3) = 1.01637\ldots\,,$$

so that $(**)$ and hence $(*)$ holds for x in J.

It follows that $f_4(x)$ is increasing for $0 < x \leqslant (\pi - \theta_4)/2$, so that in case (c) we have

$$F_M/F \leqslant 8\sqrt{3}\,\exp f_4((\pi - \theta_4)/2) = .032120\ldots\,,$$

which is smaller than the estimate found in (a). Thus the claimed result $F_M/F \leqslant (64/3)\sin^6(\pi/9)$ is established.

REFERENCES

1. H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Mathematical Library No. 19, 1967, particularly pp. 47–50.
2. Clarence Lubin, A proof of Morley's theorem, this MONTHLY 62 (1955). 110–112.
3. K. Venkatachaliengar, An elementary proof of Morley's theorem, this MONTHLY, 65 (1958) 612–613.

No other correct solutions were received.

A Card Game with a Positive Payoff

E 3330 [1989, 446]. *Proposed by Charles Vanden Eynden, Illinois State University, Normal, IL.*

In Littlewood's *A Mathematician's Miscellany* the author describes a game played with an infinite number of cards, each containing n on one side and $n + 1$ on the other, with $n = 1, 2, \ldots$. A referee chooses a card at random and holds it so that the players see opposite sides. Each player may either play or pass. If both play, the player seeing the lower number must pay the other \$1.00. Littlewood argues that the players will always pass.

This will be positive if $n_0 \geqslant 7$.

In the case of a nonrandomized strategy a clever opponent can force negative expected winnings by choosing $q_1 = 0$ and $q_n = p_{n-1}$ for $n > 1$. (A clever opponent knows the strategy of the person he is playing against.) For then, since $p_{n-1} = p_{n-1}^2 = p_{n-1}q_n$, we have

$$\begin{aligned} W &= \frac{p_1}{80} + \frac{3}{20}\sum_{n=2}^{\infty}\frac{p_n}{2^{n+1}} + \sum_{n=2}^{\infty}\frac{p_{n-1}p_{n+1}}{2^{n+1}} - \sum_{n=2}^{\infty}\frac{p_{n-1}}{2^n} \\ &= \frac{p_1}{80} + \frac{p_1p_3}{8} - \frac{p_1}{4} + \sum_{n=2}^{\infty}\left(\frac{3}{20}\frac{1}{2^{n+1}} - \frac{1}{2^{n+1}}\right)p_n + \sum_{n=2}^{\infty}\frac{p_np_{n+2}}{2^{n+2}} \\ &= p_1\left(\frac{-1.9 + p_3}{8}\right) + \sum_{n=2}^{\infty}p_n\left(\frac{-1.7 + p_{n+2}}{2^{n+2}}\right). \end{aligned}$$

Thus, since the coefficients of each p_n are negative, we have $W < 0$.

Editorial comment. Another recipe for foiling a nonrandomized strategy is to choose $q_{r+1} = 1$ if r is the smallest natural number with $p_r = 1$ and $q_i = 0$ if $i \neq r + 1$.

Solved also by P. Griffin, B. Kibler, I. Kozma (Israel), H. Lipman, and the proposer. Two incorrect solutions were received.

A Monotonic Function Connected with the Error Function

E 3351 [1989, 838]. *Proposed by Jet Wimp and Um Bencharit, Drexel University, Philadelphia, PA.*

Suppose $a > 0$, and let f be the real-valued function on $(1, \infty)$ defined by

$$f(x) = \left[1 + \operatorname{erf}(ax^{-1/2})\right]^x + \left[1 - \operatorname{erf}(ax^{-1/2})\right]^x,$$

where

$$\operatorname{erf}(u) = 2\pi^{-1/2}\int_0^u e^{-t^2}\,dt.$$

Show that f is monotonic on $(1, \infty)$.

Solution by Richard Stong, University of California at Los Angeles. We will establish that f is actually strictly increasing. Let $h(x) = \operatorname{erf}(ax^{-1/2})$, $g(x) = x$, $U(x) = (1 + h(x))^{g(x)}$, and $V(x) = (1 - h(x))^{g(x)}$. Then by elementary calculus,

$$\begin{aligned} U' &= U \cdot \left[g' \ln(1 + h) + gh'/(1 + h)\right] \\ &= (1 + h)^{g-1}\left[g'(1 + h)\ln(1 + h) + gh'\right], \\ V' &= V \cdot \left[g' \ln(1 - h) - gh'/(1 - h)\right] \\ &= (1 - h)^{g-1}\left[g'(1 - h)\ln(1 - h) - gh'\right]. \end{aligned}$$

Since $g' > 0$, we can use $\ln(1 + z) \geqslant z/(1 + z)$ and $\ln(1 - z) \geqslant -z/(1 - z)$ for $z \in [0, 1)$ to obtain $U' \geqslant (1 + h)^{g-1}[g'h + gh'] = (1 + h)^{g-1}(gh)'$ and $V' \geqslant (1 - h)^{g-1}[-g'h - gh'] = -(1 - h)^{g-1}(gh)'$. Therefore, $f' = U' + V' \geqslant [(1 + h)^{g-1} - (1 - h)^{g-1}](gh)'$. Since $h(x) \in (0, 1)$ and $g - 1 > 0$, the expression

in brackets is positive, and it suffices to show $(gh)' > 0$. By a change of variable, the expression for $\{g(x)h(x)\}'$ can be written as $(2a/\sqrt{\pi})\int_0^1 e^{-a^2t^2/x}[(1/2)x^{-1/2} + a^2t^2x^{-3/2}]\,dt$, which is clearly positive for $x > 0$.

Editorial comment. The proposers asserted the more general statement that $[1 + h(x)]^{g(x)} + [1 - h(x)]^{g(x)}$ is strictly increasing on $(1, \infty)$ as long as g, h are differentiable, $0 < h(x) < 1 < g(x)$, $g' > 0$, and $(gh)' > 0$, all on $(1, \infty)$, which can be proved as above.

They noted that the problem arises in evaluating the signal to noise ratio in the so-called polarity threshold algorithm. The fact that $f(x)$ is strictly increasing means that the signal to noise ratio increases as the number of sample points increases.

Solved also by R. A. Agnew and the proposer. Three incorrect solutions were received.

A Unique $(2k + 1)$-th Root of a Matrix?

E 3362 [1990, 63]. *Proposed by Bjorn Poonen, Winchester, MA.*

Given a positive integer n, let M denote the set of n by n matrices with complex entries. If $A \in M$, let A^* denote the conjugate transpose of A. Prove that for each positive integer k and each $B \in M$ there exists a unique $A \in M$ such that $A(A^*A)^k = B$.

Solution by Stephen Pierce, San Diego State Univ., San Diego, CA. To show existence, let $B = UH$, where U is unitary and H is positive semidefinite. Let P be the unique positive semidefinite $(2k + 1)$st root of H, so that $B = UP^{2k+1}$. If we set $A = UP$, then $A^*A = P^2$, and the result follows.

If $B = A(A^*A)^k = C(C^*C)^k$, then $B^*B = (A^*A)^{2k+1} = (C^*C)^{2k+1}$. Since B^*B is positive semidefinite, it follows that $A^*A = C^*C$. Thus the null spaces of A, C, C^*C, and A^*A are all the same. Hence A and C agree on any null vector of A^*A. If λ is a nonzero eigenvalue of $A^*A = C^*C$ corresponding to an eigenvector v, then $Bv = \lambda^k Av = \lambda^k Cv$, so that $Av = Cv$. Thus A and C agree on any eigenvector of A^*A. Since A^*A has a basis of eigenvectors, $A = C$.

Solved also by D. Callan, R. J. Chapman (U.K.), D. Ž. Đoković (Canada), M. Golomb, C. Hill, A. A. Jagers (Netherlands), F. Kittaneh (Kuwait), A. Kresch (student), O. P. Lossers (Netherlands), J.-M. Monier (France), T. S. Norfolk, B. N. Partlett, A. Pedersen (Denmark), D. Ruch, R. Stong, A. Tissier (France), University of Wyoming Problem Circle, and the proposer. Nine partially incorrect or incomplete solutions were received.

Convergence of Some Recursively Defined Series

E 3368 [1990, 150]. *Proposed by I. E. Leonard, University of Alberta, Edmonton, Alberta, Canada.*

Suppose $\alpha > 0$. Let us define a sequence $a_1, a_2, \ldots$ by taking a_1 as an arbitrary positive number and putting $a_{n+1} = a_n \exp(-a_n^\alpha)$ for $n = 1, 2, \ldots$. Find the values of α for which the following series converge:

$$\text{(i)} \quad \sum_{n=1}^{\infty} a_n, \qquad \text{(ii)} \quad \sum_{n=1}^{\infty} a_n^\alpha, \qquad \text{(iii)} \quad \sum_{n=1}^{\infty} a_n^{1-\alpha}.$$

Composite solution by Terence Shore and Douglas B. Tyler, California State University–Dominguez Hills, Carson, CA, and by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. (i) converges if and only if $\alpha < 1$, (ii) diverges for all α, and (iii) converges if and only if $0 < \alpha < 1/2$. These results follow immediately from the more general result that for all real β, the series $\sum a_n^\beta$ converges if and only if $\alpha < \beta$.

To prove this result, first note that the recurrence yields $a_n = a_1 \exp(-\sum_{j=1}^{n-1} a_j^\alpha)$ for all $n \geqslant 1$. If $\sum_{n=1}^{\infty} a_n^\alpha = S < \infty$, then $\lim a_n = a_1 \exp(-S) > 0$. This implies $\lim a_n^\alpha > 0$, which contradicts $\sum a_n^\alpha$ converging. Therefore $\sum a_n^\alpha$ must diverge, which also implies $\lim a_n = 0$.

Now suppose $\beta \leqslant \alpha$. Since $\lim a_n = 0$, we have $0 < a_n < 1$ for sufficiently large n. For such n, we have $a_n^\beta \geqslant a_n^\alpha$, and hence $\sum a_n^\beta$ diverges by the comparison test.

Now suppose $\beta > \alpha$, so that $\beta/\alpha > 1$. If we can prove $a_n^\beta < [\alpha(n-1)]^{-\beta/\alpha}$, then we obtain the convergence of $\sum a_n^\beta$ by the comparison test. To verify the desired inequality, it suffices to prove $a_n^{-\alpha} > \alpha(n-1)$ for all $n \geqslant 1$ by induction. The inequality is trivial for $n = 1$. If we assume $a_n^{-\alpha} > \alpha(n-1)$ for a particular value of n, then $a_{n+1}^{-\alpha} = a_n^{-\alpha} \exp(\alpha a_n^\alpha) > a_n^{-\alpha}(1 + \alpha a_n^\alpha) = a_n^{-\alpha} + \alpha > \alpha n$, the last step by the inductive assumption. This completes the induction and the proof.

Editorial comment. Shore and Tyler observed that on the other hand $a_n^{-\alpha} \leqslant (\alpha + 1)n + C$ for some constant C depending on a_1 and α, so that $\sum a_n^\alpha$ is comparable to the harmonic series.

Solved also by A. Fischer, E. A. Herman, P. W. Lindstrom, O. P. Lossers (Netherlands), J.-M. Monier (France), T. S. Norfolk, A. Pedersen (Denmark), A. J. Rosenthal, R. Stong, P. Tracy, University of Alberta Problem Solving Group, and the proposer. One incorrect solution was received.

Folding an Equilateral Triangle Into a Large Hexagon

E 3369 [1990, 150]. *Proposed by Jiro Fukuta, Aichi University, Aichi, Japan.*

Suppose we are given a piece of paper in the shape of an equilateral triangle ABC. Suppose P is a point in the intersection of the three open circular discs with diameters AB, BC, CA. If we fold the three corners of the paper in such a way that the vertices coincide with P, we get a hexagon three sides of which are the creases formed and the other three sides of which are portions of the sides of triangle ABC.

Prove that the area and perimeter of this hexagon are both maximized when P is the centroid of ABC.

Solution by O. P. Lossers, University of Technology, Eindhoven, The Netherlands. The hexagon η can be formed by cutting off suitable triangles $\alpha = A_1A_2A$, $\beta = B_1B_2B$, and $\gamma = C_1C_2C$ from the equilateral triangle $\tau = ABC$. Consider the triangle α; it has angles A_1, A_2, and A with $A = 60°$. By the folding structure, we have Area $\alpha = (1/8)(AP)^2(\cot A_1 + \cot A_2)$, which, by the convexity of cot on $(0, \pi/2)$, is at least

$$\frac{1}{4}(AP)^2 \cot \frac{1}{2}(A_1 + A_2) = \frac{1}{12}(AP)^2\sqrt{3}.$$

Applying the same argument to β and γ, we have

$$\text{Area}\,(\alpha \cup \beta \cup \gamma) \geqslant \frac{1}{12}\sqrt{3}\left((AP)^2 + (BP)^2 + (CP)^2\right)$$
$$\geqslant \frac{1}{12}\sqrt{3}\left((AQ)^2 + (BQ)^2 + (CQ)^2\right),$$

where Q is the centroid of τ. Hence Area $\eta \leqslant$ Area $\tau - (1/4)\sqrt{3}(AQ)^2 = (2/3)$ Area τ, with equality achieved when $P = Q$.

For the perimeter, let $\Delta_\alpha = AA_1 + AA_2 - A_1A_2$. By the folding structure, we have

$$\Delta_\alpha = \frac{1}{2}AP\left(\frac{1-\cos A_1}{\sin A_1} + \frac{1-\cos A_2}{\sin A_2}\right) = \frac{1}{2}AP\left(\tan\frac{1}{2}A_1 + \tan\frac{1}{2}A_2\right),$$

which, by the convexity of tan on $(0, \pi/2)$, is at least $AP\tan(1/4)(A_1 + A_2) = (1/3)AP\sqrt{3}$. Applying the same argument to β and γ, we have $\Delta_\alpha + \Delta_\beta + \Delta_\gamma \geqslant (\sqrt{3}/3)(AP + BP + CP) \geqslant (\sqrt{3}/3)(AQ + BQ + CQ)$, since the sum of the distances of a point from the vertices of an equilateral triangle is smallest when the point is the centroid of the triangle (cf. Ch. IX of [1] or §3.5 of [2]). Hence Perimeter η = Perimeter $\tau - (\Delta_\alpha + \Delta_\beta + \Delta_\gamma) \leqslant 3AB - \sqrt{3}AQ = 2AB$, with equality achieved when $P = Q$.

REFERENCES

1. G. Pólya, Mathematics and Plausible Reasoning, Vol. I, Princeton, 1954.
2. Nicholas D. Kazarinoff, Geometric Inequalities, New Mathematical Library, No. 4, 1961.

Editorial comment. The result for area follows from that for perimeter, since a regular hexagon has area at least equal to that of any other hexagon of equal or smaller perimeter.

Solved also by J. Dou (Spain), G. D. Chakerian and V. Schindler (Germany). Partially solved by H. Lipman and by the proposer. One incorrect solution was received.

Iterating a Polygon on the Unit Circle

E 3377 [1990, 240]. *Proposed by Farhoud Pouryoussefi (student), Sharif University of Technology, Tehran, Iran.*

Suppose we consider the polygon with vertices $e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_n}$ in the complex plane, where

$$0 \leqslant \theta_1 < \theta_2 < \cdots < \theta_n < 2\pi.$$

Suppose α and β are given positive numbers with $\alpha + \beta = 1$. Define $\theta_i^{(0)} = \theta_i$ for $i = 1, 2, \ldots, n$ and

$$\theta_i^{(k)} = \alpha\theta_i^{(k-1)} + \beta\theta_{i+1}^{(k-1)}$$

for $k = 1, 2, \ldots$, where subscripts are taken modulo n and the angles $\theta_i^{(k)}$ are taken modulo 2π. Prove that

$$\lim_{k\to\infty}\left(\theta_{i+1}^{(k)} - \theta_i^{(k)}\right) = 2\pi/n$$

for $i = 1, 2, \ldots, n$.

Editorial remark. In the first line of the problem as published there was a spurious factor 2π in the exponents.

Solution by Allan Pedersen, Søborg, Denmark. Let $\phi_i = \theta_{i+1} - \theta_i$ for $i = 1, 2, \ldots, n-1$ and $\phi_n = 2\pi + \theta_1 - \theta_n$, so that $\sum_{i=1}^{n} \phi_i = 2\pi$, $\phi_i > 0$. Define $\phi_i^{(0)} = \phi_i$ for $i = 1, 2, \ldots, n$ and

$$\phi_i^{(k)} = \alpha\phi_i^{(k-1)} + \beta\phi_{i+1}^{(k-1)} \quad (i = 1, 2, \ldots, n)$$

for $k = 1, 2, \ldots$, where the subscript $n+1$ is interpreted as 1. For any fixed k the numbers $\phi_1^{(k)}, \ldots, \phi_n^{(k)}$ are positive real numbers with sum 2π. Since $\phi_i^{(k)} = \theta_{i+1}^{(k)} - \theta_i^{(k)}$ modulo 2π, it suffices to show that $\lim_{k\to\infty} \phi_i^{(k)} = 2\pi/n$ for $i = 1, 2, \ldots, n$. Clearly

$$\left(\phi_1^{(k)}, \ldots, \phi_n^{(k)}\right)^T = M\left(\phi_1^{(k-1)}, \ldots, \phi_n^{(k-1)}\right)^T,$$

where M is the n by n matrix whose nonzero entries are $M_{i,i} = \alpha$ for $i = 1, \ldots, n$, $M_{i,i+1} = \beta$ for $i = 1, 2, \ldots, n-1$, and $M_{n,1} = \beta$. The matrix M is doubly stochastic and for $k \geqslant n-1$ the powers M^k have all entries positive. By the Perron–Frobenius theory (see D. Ruelle, *Statistical Mechanics: Rigorous Results*, Benjamin, 1969, p. 136), it follows that

$$\begin{aligned}\lim_{k\to\infty} \left(\phi_1^{(k)}, \ldots, \phi_n^{(k)}\right)^T &= \lim_{k\to\infty} M^k \left(\phi_1^{(0)}, \ldots, \phi_n^{(0)}\right)^T \\ &= \left(n^{-1} \sum_{i=1}^{n} \phi_i^{(0)}\right)(1, \ldots, 1)^T \\ &= (2\pi/n)(1, \ldots, 1)^T,\end{aligned}$$

which was to be proved. (The eigenvalues of M are $\alpha + \rho\beta$, where ρ runs over the n-th roots of unity, and so the largest eigenvalue is $\alpha + \beta = 1$, with eigenvector $(1, \ldots, 1)^T$.)

Solved also by D. Callan, R. J. Chapman (United Kingdom), R. High, H. Kappus (Switzerland), M. E. Kuczma (Poland), J. G. Merickel (student), J. H. Lindsey II, O. P. Lossers (The Netherlands), A. Müller (Switzerland), M. Riazi, R. Stong, M. Vowe (Switzerland), Western Maryland College Problems Group, and the proposer.

ADVANCED PROBLEMS

6667. *Proposed by George Baloglou and Phil Tracy, State University of New York, College at Oswego.*

If $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are positive numbers such that $a_1 a_2 \ldots a_n = b_1 b_2 \ldots b_n$ and

$$\sum_{1 \leqslant i < j \leqslant n} |a_i - a_j| \leqslant \sum_{1 \leqslant i < j \leqslant n} |b_i - b_j|,$$

(i) prove that

$$\sum_{i=1}^{n} a_i \leqslant (n-1) \sum_{i=1}^{n} b_i, \qquad \text{and}$$

(ii) show that the factor $n-1$ cannot be replaced by a smaller one.

Editorial remark. In the first line of the problem as published there was a spurious factor 2π in the exponents.

Solution by Allan Pedersen, Søborg, Denmark. Let $\phi_i = \theta_{i+1} - \theta_i$ for $i = 1, 2, \ldots, n-1$ and $\phi_n = 2\pi + \theta_1 - \theta_n$, so that $\sum_{i=1}^{n} \phi_i = 2\pi$, $\phi_i > 0$. Define $\phi_i^{(0)} = \phi_i$ for $i = 1, 2, \ldots, n$ and

$$\phi_i^{(k)} = \alpha\phi_i^{(k-1)} + \beta\phi_{i+1}^{(k-1)} \quad (i = 1, 2, \ldots, n)$$

for $k = 1, 2, \ldots$, where the subscript $n+1$ is interpreted as 1. For any fixed k the numbers $\phi_1^{(k)}, \ldots, \phi_n^{(k)}$ are positive real numbers with sum 2π. Since $\phi_i^{(k)} = \theta_{i+1}^{(k)} - \theta_i^{(k)}$ modulo 2π, it suffices to show that $\lim_{k\to\infty} \phi_i^{(k)} = 2\pi/n$ for $i = 1, 2, \ldots, n$. Clearly

$$\left(\phi_1^{(k)}, \ldots, \phi_n^{(k)}\right)^T = M\left(\phi_1^{(k-1)}, \ldots, \phi_n^{(k-1)}\right)^T,$$

where M is the n by n matrix whose nonzero entries are $M_{i,i} = \alpha$ for $i = 1, \ldots, n$, $M_{i,i+1} = \beta$ for $i = 1, 2, \ldots, n-1$, and $M_{n,1} = \beta$. The matrix M is doubly stochastic and for $k \geqslant n-1$ the powers M^k have all entries positive. By the Perron–Frobenius theory (see D. Ruelle, *Statistical Mechanics: Rigorous Results*, Benjamin, 1969, p. 136), it follows that

$$\begin{aligned}\lim_{k\to\infty} \left(\phi_1^{(k)}, \ldots, \phi_n^{(k)}\right)^T &= \lim_{k\to\infty} M^k \left(\phi_1^{(0)}, \ldots, \phi_n^{(0)}\right)^T \\ &= \left(n^{-1} \sum_{i=1}^{n} \phi_i^{(0)}\right)(1, \ldots, 1)^T \\ &= (2\pi/n)(1, \ldots, 1)^T,\end{aligned}$$

which was to be proved. (The eigenvalues of M are $\alpha + \rho\beta$, where ρ runs over the n-th roots of unity, and so the largest eigenvalue is $\alpha + \beta = 1$, with eigenvector $(1, \ldots, 1)^T$.)

Solved also by D. Callan, R. J. Chapman (United Kingdom), R. High, H. Kappus (Switzerland), M. E. Kuczma (Poland), J. G. Merickel (student), J. H. Lindsey II, O. P. Lossers (The Netherlands), A. Müller (Switzerland), M. Riazi, R. Stong, M. Vowe (Switzerland), Western Maryland College Problems Group, and the proposer.

ADVANCED PROBLEMS

6667. *Proposed by George Baloglou and Phil Tracy, State University of New York, College at Oswego.*

If $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are positive numbers such that $a_1 a_2 \ldots a_n = b_1 b_2 \ldots b_n$ and

$$\sum_{1 \leqslant i < j \leqslant n} |a_i - a_j| \leqslant \sum_{1 \leqslant i < j \leqslant n} |b_i - b_j|,$$

(i) prove that

$$\sum_{i=1}^{n} a_i \leqslant (n-1) \sum_{i=1}^{n} b_i, \qquad \text{and}$$

(ii) show that the factor $n-1$ cannot be replaced by a smaller one.

6668. *Proposed by Dragomir Ž. Đoković, University of Waterloo, Ontario, Canada.*

If n is even, show that there exists $A \in \mathrm{GL}_n(\mathbb{C})$ which cannot be represented in the form $A = QS$ with $Q \in \mathrm{SO}_n(\mathbb{C})$ and S symmetric. (For every n it is known that every $A \in \mathrm{GL}_n(\mathbb{C})$ can be represented as $A = QS$ with $Q \in \mathrm{O}_n(\mathbb{C})$ and S symmetric.)

6669. *Proposed by Paul Erdős, Hungarian Academy of Sciences, Budapest.*

Prove that there is a constant c such that if $n! = a!b!$ with $1 < a < b < n$, then

$$n - b < c \log\log n.$$

(It has been conjectured, but never proved, that if $n > 10$ and $n! = a!b!$ with $1 < a < b < n$, then $b = n - 1$ and $n = a!$. For $n = 10$ there is the exception $10! = 6!7!$.)

SOLUTIONS OF ADVANCED PROBLEMS

The Sum of the Prime Factors of a Positive Integer

6599 [1989, 366]. *Proposed by Nick MacKinnon, Winchester College, Winchester, Great Britain.*

Let $m(n)$ denote the sum of the prime factors of n, repetitions counting. For example $m(12) = 2 + 2 + 3 = 7$. What is the asymptotic behavior of

$$\sum_{n=1}^{N} m(n)$$

for large N?

Solution by Aleksandar Ivić, University of Belgrade, Yugoslavia. For any fixed integer $M \geqslant 1$ we have

$$\sum_{n=1}^{N} m(n) = \frac{d_1 N^2}{\log N} + \frac{d_2 N^2}{\log^2 N} + \cdots + \frac{d_M N^2}{\log^M N} + O\left(\frac{N^2}{\log^{M+1} N}\right), \tag{1}$$

where $d_1 = \pi^2/12$ and $d_2, \ldots, d_M$ are suitable constants. Formula (1) is proved in [2], where further references are given; a simpler proof is given in [3]. Here is a brief version of the proof.

Let $\beta(n)$ denote the sum of the *distinct* prime factors of n. If p runs over the primes and i, j over positive integers, we have

$$\begin{aligned}
\sum_{n \leqslant N} m(n) - \sum_{n \leqslant N} \beta(n) &= \sum_{n \leqslant N} \sum_{p^i \mid n,\, p^{i+1} \nmid n} (i-1)p \\
&= \sum_{p^j \leqslant N,\, j \geqslant 2} p\lfloor Np^{-j} \rfloor \\
&= N \sum_{p \leqslant N^{1/2}} p^{-1} + O\left(\sum_{p \leqslant N^{1/2}} p\right) + O\left(\sum_{p \leqslant N^{1/3}} Np^{-2}\right) \\
&= N \log\log N + O(N).
\end{aligned}$$

(A sharper formula for $\Sigma m(n) - \Sigma\beta(n)$ may be found in **[4]**.) Thus it suffices to prove (1) with m replaced by β.

Now let $S(x) = \Sigma_{p \leqslant x} p$. Then if $\pi(x)$ denotes the number of primes not exceeding x and

$$\Delta(x) = \pi(x) - \int_2^x (\log t)^{-1}\, dt,$$

the estimate $\Delta(x) = O(x(\log x)^{-M-1})$ of de la Vallée Poussin gives

$$\begin{aligned} S(x) &= \int_2^x \frac{t\,dt}{\log t} + \int_2^x t\,d\Delta(t) \\ &= \int_2^x \frac{t\,dt}{\log t} + x\Delta(x) - \int_2^x \Delta(t)\,dt \\ &= \int_2^x \frac{t\,dt}{\log t} + O\left(\frac{x^2}{(\log x)^{M+1}}\right). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{n \leqslant N} \beta(n) &= \sum_{pm \leqslant N} p = \sum_{m \leqslant N/2} S(N/m) \\ &= \sum_{m \leqslant N/2} \left\{ \int_2^{N/m} \frac{t\,dt}{\log t} + O\left(\frac{N^2}{m^2 \log^{M+1}(N/m)}\right)\right\} \\ &= N^2 \sum_{m \leqslant N/2} \int_m^{N/2} \frac{du}{u^3 \log(N/u)} + O\left(\frac{N^2}{\log^{M+1} N}\right), \end{aligned}$$

since

$$\sum_{m \leqslant N/2} \frac{N^2}{m^2 \log^{M+1}(N/m)} = \sum_{m \leqslant N^{1/2}} + \sum_{N^{1/2} < m \leqslant N/2} = O\left(\frac{N^2}{\log^{M+1} N}\right).$$

Thus

$$\sum_{n \leqslant N} \beta(n) = N^2 \int_1^{N/2} \frac{\lfloor u \rfloor\, du}{u^3 \log(N/u)} + O\left(\frac{N^2}{\log^{M+1} N}\right).$$

If we now use

$$\frac{1}{\log(N/u)} = \frac{1}{\log N}\left\{ \sum_{k=0}^{M-1} \left(\frac{\log u}{\log N}\right)^k + O\left(\left(\frac{\log u}{\log N}\right)^M\right)\right\},$$

the desired result follows on writing $\int_1^{N/2} = \int_1^{\infty} - \int_{N/2}^{\infty}$ and noting that

$$\int_1^{\infty} \frac{\lfloor u \rfloor\, du}{u^3} = \frac{1}{2}\int_{1-}^{\infty} \frac{d\lfloor u \rfloor}{u^2} = \frac{1}{2}\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{12}.$$

Editorial comment. The assertion

$$\sum_{n=1}^{N} m(n) = \pi^2 N^2/(12 \log N) + O(N^2/\log^2 N)$$

given by the special case $M = 1$ of (1) occurs as formula (6.13) on page 150 of **[1]**.

REFERENCES

1. J.-M. De Koninck and A. Ivić, Topics in Arithmetical Functions, North-Holland Publishing Co., 1980.
2. _____, The distribution of the average prime divisor of an integer, *Archiv. Math.*, 43 (1984) 37–43.
3. J.-M. De Koninck and R. Sitaramachandrarao, Sums involving the largest prime divisor of an integer II, *Indian J. Pure Appl. Math.*, 19 (1988) 990–1004.
4. A. Ivić, The distribution of quotients of small and large additive functions, *Bolletino Unione Mat. Ital.*, (7) 2-B (1988) 79–97.

Solved also by Jean Anglesio (France), Joachim Herzog (Germany), Kee-Wai Lau (Hong Kong), Enoch K. S. Lee, Reiner Martin, Thomas Maxsein & Paul R. Smith (Germany), Ondrej Šuch (Czechoslovakia), and two referees. Two incomplete solutions were received.

A Nondecomposition Theorem

6620 [1990, 76]. *Proposed by W. Allegretto and A. Meir, University of Alberta, Edmonton, Canada.*

Suppose f is locally integrable on R, but

$$\int_R |f(x)|\,dx = \infty.$$

Show that for every $p > 0$ there exists $\phi \in L^p(R)$ such that it is impossible to have

$$f(x)\phi(x) = g(x)h(x) \qquad (x \in R),$$

where $g \in L^1(R)$ and $h \in L^p(R)$.

Solution by Adam Riese, Wright State University, Dayton, OH. It is enough to do this for non-negative f that vanish for $x < 0$.

Define k_n inductively by $k_0 = 0$ and

$$\int_{k_{n-1}}^{k_n} f(t)\,dt = 1 \qquad (n = 1, 2, \ldots).$$

Let $\epsilon = p/2$ and for $n \geqslant 1$ let

$$r(x) = \frac{1}{n}, \qquad k_{n-1} \leqslant x < k_n.$$

We now show that $\phi(x) = r(x)^{(1+\epsilon)/p} f(x)^{1/p}$ is a function of the desired sort. Since

$$\int_0^\infty \phi(x)^p\,dx = \int_0^\infty r(x)^{1+\epsilon} f(x)\,dx = \sum_{n=1}^{\infty} \int_{k_{n-1}}^{k_n} \frac{1}{n^{1+\epsilon}} f(x)\,dx < \infty,$$

it is in L^p. Now assume $f(x)\phi(x) = g(x)h(x)$, where $g \in L^1$ and $h \in L^p$. Let A be the set of all $x > 0$ on which $g(x) \geqslant r(x)f(x)$, and let B be the complement of this set. For x in B we have $h(x) \geqslant \phi(x)/r(x)$, so that

$$h(x)^p \geqslant r(x)^{1+\epsilon-p} f(x) \geqslant r(x)f(x).$$

But

$$\int_A g(x)\,dx + \int_B h(x)^p\,dx \geqslant \int_0^\infty r(x)f(x)\,dx = \sum_{n=1}^{\infty} \frac{1}{n},$$

a contradiction to $g \in L^1$ and $h \in L^p$.

Solved also by Jaime E. Muñoz Rivera (Brazil), Kenneth Schilling, and the proposers.

No Such Function Exists

6622 [1990, 161]. *Proposed by Robert C. Dalang, École Polytechnique Fédérale, Lausanne, Switzerland.*

Does there exist a Lebesgue measurable function $f\colon [0,1] \to [0,1]$ which is positive on a set of positive measure and which satisfies

$$\sum_{n=1}^{\infty} \sum_{j=0}^{2^{n-1}-1} \left\{ \int_{2j/2^n}^{(2j+1)/2^n} f(x)\,dx \int_{(2j+1)/2^n}^{(2j+2)/2^n} f(x)\,dx \right\}^{1/2} < +\infty?$$

Solution by Marcin E. Kuczma, Institute of Mathematics, University of Warsaw, Poland. No. Assume such an f exists. Clearly $f(x) \geqslant c > 0$ on a set S of positive measure. There is an interval I (in fact, one of the form $[p2^{-q}, (p+1)2^{-q}]$ with positive integers p and q) for which the measure inequality

$$m(I \cap S) \geqslant \tfrac{7}{8} m(I) \tag{1}$$

is valid. This follows from the Lebesgue density theorem (or more elementarily, from the outer regularity and translation invariance of Lebesgue measure).

Fix $n > q$ and partition I into 2^{n-q-1} equal nonoverlapping intervals I_i each of length 2^{-n+1} $(i = 1, \ldots, 2^{n-q-1})$. We easily see that at least half of them satisfy

$$m(I_i \cap S) \geqslant \tfrac{3}{4} m(I_i). \tag{2}$$

Otherwise,

$$m(I_i \setminus S) > m(I_i)/4$$

would hold for more than half of them, and by summation

$$m(I \setminus S) > m(I)/8,$$

contrary to (1).

Hence we may choose 2^{n-q-2} intervals from the I_i each satisfying (2); call them J_i $(i = 1, \ldots, 2^{n-q-2})$. Now subdivide each J_i into two equal halves H_i and K_i. Then

$$\begin{aligned} m(H_i \cap S) &\geqslant m(J_i \cap S) - m(K_i) \\ &\geqslant \tfrac{3}{4} m(J_i) - \tfrac{1}{2} m(J_i) = 2^{-n-1}, \end{aligned}$$

and similarly for $m(K_i \cap S)$. Hence

$$\int_{H_i} f \int_{K_i} f \geqslant c^2 2^{-2n-2}, \qquad 1 \leqslant i \leqslant 2^{n-q-2},$$

and the sum of the square roots of these products is at least

$$2^{-n-1} 2^{n-q-2} c = 2^{-q-3} c > 0.$$

Thus all terms of the outer series with $n \geqslant q+1$ exceed a fixed positive constant.

Solved also by Duane W. Bailey, Richard A. Stong, and the proposer.

Riemann Integrability and Continuity

6626 [1990, 252]. *Proposed by Bruce G. Dearden and Michael B. Gregory, University of North Dakota, Grand Forks, ND.*

It is well known that if ϕ is a continuous function on $\mathbb{R}$, then the composition $\phi \circ f$ is Riemann integrable on $[a, b]$ whenever f is Riemann integrable on $[a, b]$. Is the converse true? That is, suppose $\phi\colon \mathbb{R} \Rightarrow \mathbb{R}$ has the property that $\phi \circ f$ is Riemann integrable on $[a, b]$ whenever f is Riemann integrable on $[a, b]$. Does it follow that ϕ is continuous?

Solution I by James R. Munkres, MIT, Cambridge, MA. Yes. This is a consequence of a stronger result.

THEOREM. *Let $\phi\colon \mathbb{R} \to \mathbb{R}$ have the property that $\phi \circ f$ is Riemann-integrable on $[a, b]$ whenever f is continuous on $[a, b]$. Then ϕ is continuous.*

Proof. Let C be a closed subset of $[a, b]$ that contains no non-trivial open interval but has positive measure (see, e.g., Problem 3-11 of **[2]**; such sets are sometimes called "fat Cantor sets"). Let $g\colon \mathbb{R} \to \mathbb{R}$ be a continuous function that is positive on $(a, b) \setminus C$ and zero elsewhere. Given $p \in \mathbb{R}$ define another continuous function by

$$f(x) = p + g(x).$$

By Lebesgue's theorem (f is Riemann integrable if and only if f is bounded and the set of points at which f is discontinuous has Lebesgue measure zero) the function $\phi \circ f$ must be continuous at some point c of (a, b) belonging to C. We claim that ϕ is continuous from the right at p. Given $\epsilon > 0$, choose $\delta > 0$ such that

$$|\phi(f(x)) - \phi(f(c))| < \epsilon \quad \text{for} \quad |x - c| < \delta.$$

Choose d so $0 < d - c < \delta$ and $g(d) > 0$. For

$$y \in [f(c), f(d)] = [p, f(d)] = [p, p + g(d)]$$

there is an $x \in [c, d]$ with $f(x) = y$. Hence,

$$|\phi(y) - \phi(p)| = |\phi(y) - \phi(p + g(c))| = |\phi(f(x)) - \phi(f(c))| < \epsilon.$$

Since a similar argument with $f(x) = p - g(x)$ shows that ϕ is continuous from the left at p, the result follows.

The theorem is actually valid with the hypothesis "f is of class C^∞." To prove this fact, replace the $g(x)$ above by a C^∞ function defined as follows. Write the complement of C relative to (a, b) as the union of a countable number of open intervals U_i. Let g_i be a C^∞ function positive on U_i and zero elsewhere. Let M_i be the maximum of the absolute values of g_i and its derivatives of order at most i. Finally, set

$$g = \sum_i \frac{g_i}{2^i M_i}.$$

Solution II by Lee Mohler, St. Martin's College, Lacey, WA. Yes; we prove the contrapositive. Suppose φ is discontinuous at some $x_0 \in \mathbb{R}$. We will construct a

Riemann integrable function f on the closed interval $[0, 1]$ such that $\varphi \circ f$ is not Riemann integrable. Moreover, a somewhat more elaborate construction provides a *continuous* such f.

There is a real sequence $\{x_i\}$ such that $x_i \to x_0$ but $\varphi(x_i) \to y$ where $y \neq \varphi(x_0)$; the possibility that $y = \pm\infty$ does not change the argument. Let $q_1, q_2, \ldots$ be an enumeration of the set $\mathbb{Q}$ of rational numbers in $[0, 1]$. Let P be the set of irrationals. Define f on $[0, 1]$ by setting $f(q_i) = x_i$ and $f(x) = x_0$ for all x in P. It is straightforward to check that f is continuous on P. (This f is similar to the well-known example of a function continuous on P and discontinuous on $\mathbb{Q}$. See, e.g. [**1**, p. 27].) The discontinuities of f are countable, so f is Riemann integrable.

Given x in P there is a subsequence $\{q_{v(i)}\}$ of $\{q_i\}$ that converges to x. Observe that

$$\lim_{i\to\infty} \varphi \circ f(q_{v(i)}) = \lim_{i\to\infty} \varphi(x_{v(i)}) = y \neq \varphi(x_0) = \varphi \circ f(x).$$

Hence $\varphi \circ f$ is discontinuous on P and so is not Riemann integrable.

We now sketch a somewhat more elaborate construction that provides a continuous f. Recall that the classical Cantor set in $[0, 1]$ is obtained by recursively deleting a sequence of "open middle thirds" from a sequence of closed intervals starting with $[0, 1]$. If somewhat smaller open intervals are removed at each stage of the construction, we obtain a set homeomorphic to the Cantor set, but with positive Lebesgue measure [**1**, p. 88]. Let P_0 denote this set, and Q_0 the set of midpoints of the deleted intervals. Let $q_1, q_2, \ldots$ be an enumeration of Q_0. Define f as above on the set $P_0 \cup Q_0$. Just as before, f is continuous on P_0. Since each point of Q_0 is isolated in $P_0 \cup Q_0$, the function f is continuous on Q_0. Moreover, $P_0 \cup Q_0$ is closed. Hence by the Tietze extension theorem, f can be extended continuously to all of $[0, 1]$. But again as before, $\varphi \circ f$ is discontinuous at every point of P_0, a set with positive Lebesgue measure. Therefore $\varphi \circ f$ is not Riemann integrable.

REFERENCES

1. B. Gelbaum and J. Olmsted, *Counterexamples in Analysis*, Holden-Day, 1964.
2. M. Spivak, *Calculus on Manifolds*, Benjamin, 1965.

Solved also by William C. Allen, D. R. Arterburn, Duane W. Bailey, Michael W. Botsko, Central Michigan University Problem Group, Frank J. Curtis, Lawrence E. Flynn, Michael B. Gregory, Don Hadwin, Chris Hill, Marcin E. Kuczma (Poland), Francisco García Mazarío (Spain), James G. Merickel, Mark D. Meyerson, Harry I. Miller (Yugoslavia), Jean-Marie Monier (France), Howard and Vicky Morris, Adam Riese, Kenneth A. Ross, Günter Rote (Austria), Kenneth Schilling, Richard Stong, Dale Varberg, Xianggen Xia, and the proposers. One incorrect solution was received.

Triangles With Integer Sides Whose Area Is the Square of an Integer

6628 [1990, 350]. *Proposed by Robert A. Melter, Long Island University, Southampton, NY.*

Call a triangle a Heron Triangle if it has integer sides and integer area. Fermat showed that there does not exist a Heron right triangle whose area is a perfect square. However, the triangle with sides 9, 10, 17 has area 36.

Prove that there are infinitely many Heron triangles whose sides have no common factor and whose area is a perfect square.

Riemann integrable function f on the closed interval $[0, 1]$ such that $\varphi \circ f$ is not Riemann integrable. Moreover, a somewhat more elaborate construction provides a *continuous* such f.

There is a real sequence $\{x_i\}$ such that $x_i \to x_0$ but $\varphi(x_i) \to y$ where $y \neq \varphi(x_0)$; the possibility that $y = \pm\infty$ does not change the argument. Let $q_1, q_2, \ldots$ be an enumeration of the set $\mathbb{Q}$ of rational numbers in $[0, 1]$. Let P be the set of irrationals. Define f on $[0, 1]$ by setting $f(q_i) = x_i$ and $f(x) = x_0$ for all x in P. It is straightforward to check that f is continuous on P. (This f is similar to the well-known example of a function continuous on P and discontinuous on $\mathbb{Q}$. See, e.g. [**1**, p. 27].) The discontinuities of f are countable, so f is Riemann integrable.

Given x in P there is a subsequence $\{q_{v(i)}\}$ of $\{q_i\}$ that converges to x. Observe that

$$\lim_{i\to\infty} \varphi \circ f(q_{v(i)}) = \lim_{i\to\infty} \varphi(x_{v(i)}) = y \neq \varphi(x_0) = \varphi \circ f(x).$$

Hence $\varphi \circ f$ is discontinuous on P and so is not Riemann integrable.

We now sketch a somewhat more elaborate construction that provides a continuous f. Recall that the classical Cantor set in $[0, 1]$ is obtained by recursively deleting a sequence of "open middle thirds" from a sequence of closed intervals starting with $[0, 1]$. If somewhat smaller open intervals are removed at each stage of the construction, we obtain a set homeomorphic to the Cantor set, but with positive Lebesgue measure [**1**, p. 88]. Let P_0 denote this set, and Q_0 the set of midpoints of the deleted intervals. Let $q_1, q_2, \ldots$ be an enumeration of Q_0. Define f as above on the set $P_0 \cup Q_0$. Just as before, f is continuous on P_0. Since each point of Q_0 is isolated in $P_0 \cup Q_0$, the function f is continuous on Q_0. Moreover, $P_0 \cup Q_0$ is closed. Hence by the Tietze extension theorem, f can be extended continuously to all of $[0, 1]$. But again as before, $\varphi \circ f$ is discontinuous at every point of P_0, a set with positive Lebesgue measure. Therefore $\varphi \circ f$ is not Riemann integrable.

REFERENCES

1. B. Gelbaum and J. Olmsted, *Counterexamples in Analysis*, Holden-Day, 1964.
2. M. Spivak, *Calculus on Manifolds*, Benjamin, 1965.

Solved also by William C. Allen, D. R. Arterburn, Duane W. Bailey, Michael W. Botsko, Central Michigan University Problem Group, Frank J. Curtis, Lawrence E. Flynn, Michael B. Gregory, Don Hadwin, Chris Hill, Marcin E. Kuczma (Poland), Francisco García Mazarío (Spain), James G. Merickel, Mark D. Meyerson, Harry I. Miller (Yugoslavia), Jean-Marie Monier (France), Howard and Vicky Morris, Adam Riese, Kenneth A. Ross, Günter Rote (Austria), Kenneth Schilling, Richard Stong, Dale Varberg, Xianggen Xia, and the proposers. One incorrect solution was received.

Triangles With Integer Sides Whose Area Is the Square of an Integer

6628 [1990, 350]. *Proposed by Robert A. Melter, Long Island University, Southampton, NY.*

Call a triangle a Heron Triangle if it has integer sides and integer area. Fermat showed that there does not exist a Heron right triangle whose area is a perfect square. However, the triangle with sides 9, 10, 17 has area 36.

Prove that there are infinitely many Heron triangles whose sides have no common factor and whose area is a perfect square.

Solution I by Charles R. Maderer, Indiana University of Pennsylvania, Indiana, PA. For each positive integer k define

$$a(k) = 20k^4 + 4k^2 + 1, \qquad b(k) = 8k^6 - 4k^4 - 2k^2 + 1,$$
$$c(k) = 8k^6 + 8k^4 + 10k^2.$$

Clearly

$$\max\{a(k), b(k)\} < c(k) < a(k) + b(k),$$

so that $a(k)$, $b(k)$, and $c(k)$ are the sides of a triangle with longest side $c(k)$. By Heron's formula this triangle has area

$$\{(2k)(2k^2 - 1)(2k^2 + 1)\}^2.$$

Since

$$a(k) + b(k) - c(k) = 2(2k^2 - 1)^2, \qquad a(k) + b(k) + c(k) = 2(2k^2 + 1)^3,$$

it is immediate that $a(k)$, $b(k)$, and $c(k)$ have no common factor greater than one. The assertion follows.

Solution II by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. An infinite family of such Heron triangles may be constructed inductively. Define $a_1 = 9$ and $a_{n+1} = (2a_n - 1)^2$ for $n \geq 1$. Since $2a_1 - 2 = 4^2$ and

$$2a_{n+1} - 2 = 4a_n(2a_n - 2),$$

it is clear that both a_n and $2a_n - 2$ are perfect squares for all $n \geq 1$. Now the numbers

$$a_{n+1}, (a_n - 1)a_{n+1} + 1, \qquad a_n a_{n+1} - 1$$

are easily seen to have no common factor, and to be the side lengths of a triangle. This triangle has semiperimeter $a_n a_{n+1}$, so by Heron's formula its area is

$$\sqrt{a_n a_{n+1} \cdot (a_n - 1)a_{n+1} \cdot (a_{n+1} - 1) \cdot 1} = a_{n+1}a_n(2a_n - 2),$$

a perfect square.

Solution III by Jim Buddenhagen, Southwestern Bell Telephone Co., St. Louis, MO. We show that there are infinitely many pairs of Heron triangles which share the same square area. To see this suppose m is an odd positive integer greater than 1 such that $(m^2 - 1)/2$ is a perfect square. From the theory of the Pell equation this means that m has the form

$$m = (3 + 2\sqrt{2})^j/2 + (3 - 2\sqrt{2})^j/2$$

for some positive integer j. Then the triangle with sides

$$(m^3 + m^2)/2 - 1, (m^3 - m^2)/2 + 1, m^2$$

and the triangle with sides

$$m^3 - (m - 1)/2, m^3 - (m + 1)/2, m$$

both have area $m^2(m^2 - 1)/2$. For example if $j = 1$ and $m = 3$, the Heron triangle with sides 3, 25, 26 has the same area 36 as the Heron triangle with sides 9, 10, 17.

It is not clear if there is an absolute bound on the number of Heron triangles that can share the same square area; the triangles with sides (113, 3137, 3150), (800, 1241, 2009), and (245, 1443, 1448) all have area 420^2.

Editorial comment. In the article *On* $A^4 + B^4 + C^4 = D^4$, *Math. Comp.*, 51 (1988) 825–835, Noam Elkies proved that there are infinitely many quadruples of positive integers A, B, C, D such that

$$\gcd(A, B, C, D) = 1, \qquad A^4 + B^4 + C^4 = D^4.$$

Since fourth powers are congruent to 0 or 1 modulo 16, in such quadruples exactly one of A, B, C is odd. Several solvers remarked that if we choose

$$a = B^4 + C^4, \qquad b = A^4 + C^4, \qquad c = A^4 + B^4,$$

then $\gcd(a, b, c) = 1$ and a, b, and c form a triangle with area $(ABCD)^2$. This was the only solution of which the editors were aware when the problem was published.

Several solvers used the theory of elliptic curves (cf. §5.7 of I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory of Numbers*, Wiley, 1991 or J. H. Silverman, *The Arithmetic of Elliptic Curves*, Springer, 1986). For example, Jim Buddenhagen noted that a Heron triangle whose sides have no common factor is uniquely determined by its angles, or equivalently by the tangents t_1, t_2, t_3 of its half-angles. The formulas

$$t_1 = F/\{s(s-a)\}, \qquad t_2 = F/\{s(s-b)\}, \qquad t_3 = F/\{s(s-c)\},$$

where F is the area and s is the semi-perimeter of the triangle, show that for a Heron triangle t_1, t_2, t_3 are rational numbers. For example, the triangle with sides 9, 10, 17 and area 36 has $(t_1, t_2, t_3) = (2/9, 1/4, 2)$. The formula $t_1 t_2 t_3 = F/s^2$ shows that the area of a Heron triangle is a perfect square if and only if $t_1 t_2 t_3$ is the square of a rational number. Since the half-angles of a triangle add up to $\pi/2$ and thus $t_3 = (1 - t_1 t_2)/(t_1 + t_2)$, we are led to the family of elliptic curves

$$u^2 = t_1 t_2 (t_1 + t_2)(1 - t_1 t_2), \qquad (*)$$

where t_2, say, is a parameter and t_1 and u are variables. We need not be concerned whether t_1 is positive or negative, since if (t_1, u) is on $(*)$ so is $(-1/t_1, u/t_1^2)$. To complete the solution by this method we need only show that for some particular value of t_2, for example $t_2 = 1/4$, the elliptic curve $(*)$ has positive rank.

Nathan J. Fine pointed out that formula (26) of his paper, "On rational triangles," this MONTHLY, 83 (1976) 517–521, provides a solution along the same lines as Solution I.

Solved also by Kevin Brown and Daniel Dufresne, Nathan J. Fine, Lorraine L. Foster, Richard K. Guy (Canada), Friend H. Kierstead, Jr., L. E. Mattics, Les Reid, David G. Robinson, Raul A. Simon (Chile), and Ivan Vidav (Yugoslavia). One anonymous solution was also received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

How Does One Cut A Triangle? By Alexander Soifer, Center for Excellence in Mathematical Education, 885 Red Mesa Drive, Colorado Springs, CO 80906, 1990, xiii + 138 pp.

MURRAY S. KLAMKIN
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

It is some thirty years since Dieudonné's Royaumont Seminar Address [**1**] in which he proclaimed "Euclid Must Go" and that it should be replaced by two-dimensional linear algebra. Dieudonné's quarrel was not with the purpose of geometry, which he considered important, but with the *method* of teaching geometry. Dieudonné's reasons were that Euclid as taught was a "Process Fantastically Laborious" as well as having "the affine and metric properties hopelessly mixed up." While there are many who would agree with the latter reasons, there are many who do not agree with the linear algebra replacement as proposed. For a good late survey of this, see Burn [**2**] and references therein. Burn notes that Freudenthal [**3**] in his chapter on "The Case of Geometry" offers a profound critique of Dieudonné and recommends a piecemeal school geometry program with strong intuitive spatial content. As to "What is Geometry?," see Freudenthal [**3**], Chern [**4, 5**], Yaglom [**6**] and Kelly [**7**].

Although Dieudonné's message was not heeded in the U.S. and Canada, it now appears that in our secondary schools Euclid and other geometry is essentially gone and has not been replaced. This treatment of geometry in our schools should be contrasted with that in the U.S.S.R., Hungary and Romania, for example, in which geometry is covered repeatedly in the different grades. They have many good books with challenging problems, e.g., [**8–17**] (fortunately, these have been translated into English) unlike most of the sterile ones used in the U.S. and Canada which take great pains to prove very obvious theorems but not unobvious attractive theorems and insist that proofs be done in "two columns."†

Some time ago Poincaré addressed the downgrading of geometry in his classic book *The Foundation of Science* [**18**] which I strongly believe should be required reading for all mathematical educators and teachers. I quote:

> It looks as if geometry could contain nothing which is not already included in algebra or analysis; that geometrical facts are only algebraic or analytic facts expressed in another language. It might then be thought that after our review there would remain nothing more for us to say relating specially to geometry. This would be to fail to recognize the importance of well-constructed language, not to comprehend what is added to the things themselves by the method of expressing these things and consequently of grouping them.
>
> First the geometric considerations lead us to set ourselves new problems; these may be, if you choose, analytic problems, but such as we never would have set ourselves in connection with

†Apparently in the 1990 geometry syllabus in Alberta, all proofs have been removed.

analysis. Analysis profits by them however, as it profits by those it has to solve and satisfy the needs of physics.

A great advantage of geometry lies in the fact that in it the senses can come to the aid of thought, and help find the path to follow, and many minds prefer to put the problems of analysis into geometric form. Unhappily our senses can not carry us very far, and they desert us when we wish to soar beyond the classical three dimensions. Does this mean that, beyond the restricted domain wherein they seem to wish to imprison us, we should rely only on pure analysis and that all geometry of more than three dimensions is vain and object-less? The greatest masters of a preceding generation would have answered 'yes'; today we are so familiarized with this notion that we can speak of it, even in a university course, without arousing too much astonishment.

But what good is it? That is easy to see: First it gives us a very convenient terminology, which expresses concisely what the ordinary analytic language would say in prolix phrases. Moreover, this language makes us call like things by the same name and emphasize analogies it will never again let us forget. It enables us therefore still to find our way in this space which is too big for us and which we can not see, always recalling visible space, which is only an imperfect image of it doubtless, but which is nevertheless an image. Here, again, as in all the preceding examples, it is analogy with the simple which enables us to comprehend the complex.

The downgrading of geometry also affects many courses down the line. In the last few years, there has been lots of activity about improving our calculus courses [**19–20**]. No doubt there are improvements to be made in our textbooks and teaching methods but another improvement should be in the students' secondary school preparation. In particular, I and a number of my colleagues have found that wherever there is any geometry involved in the calculus, and there is quite a bit, students have all kinds of difficulties. In this regard, I quote from the beginning of the late N. Steenrod's first lecture at a geometry conference [**21**] which addresses this point and which is highly in tune with Pioncaré (above).

In this series of lectures I am going to discuss the geometric content of the freshman and sophomore mathematics courses. I shall criticize what we as teachers are now doing and suggest what we might do. Let me begin with what I believe to be the chief criticisms.

(1) Although geometry pervades all of mathematics and is present at every stage of a development, too often do we fail to point this out to our students. We rely on analytical formulations since we realize that they are complete and we are in a hurry to get on to other material. We do not take time to look at geometric formulations.

(2) We are too greatly impressed by the rigor of analysis. We seem to feel that geometry is not rigorous, or at least that the background needed for rigor is not avaialble. We feel that it is better not to do anything that is not rigorous. I think we are buffaloed too much by this.

(3) When we do present geometry, it is too often the instructor who does the geometry while the student is merely a passive recipient. We present the geometry to him in order to explain the analysis, but then we require him to do only the analysis—no geometry.

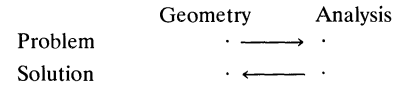
(4) We tend to avoid geometric formulations of questions in examinations. Questions are difficult to formulate geometrically. Almost any time you try such a question, you find that a large group of students misinterpret it. Such questions are hard to grade because the answers are so varied. The absence of geometric questions on final exams tends to degrade the geometric content of the course, and leads to its neglect.

Now that I have listed the main criticisms, let me take them up one at a time and fill in some details. The pervasiveness of geometry is an idea that goes back to Descartes, for a coordinate system in the plane or in space sets up an equivalence between geometry and algebra-analysis. Every geometric proposition can be translated into its algebraic-analytic analog and vice versa. I am not proposing that we lead the student through the details of the formal isomorphism between these two systems, but I am trying to remind you that the geometry is always there and to keep in mind that the geometric language for the conversion is always at hand. For example,

here are a few geometric terms and their algebraic-analytic counterparts:

Geometric language	*Algebraic-Analytic language*
point, vector	number triple (x, y, z)
projection	coordinate, variable
surface	equation
plane	linear equation
region	system of inequalities
mapping, transformation	function
neighborhood	ε, δ
limit (using deleted neighborhoods)	limit (using e, δ)
tangent	derivative

One might ask, in view of this equivalence, why bother with the geometry at all? The answer is clear to all of us. The first main reason is that many applications of calculus are to problems of conics, oscillating systems, the 2-body problem. A complete solution of such a problem has three main steps: The reformulation in analytic terms, the derivation of an analytic solution, and the interpretation of the solution in geometric terms.



A second reason is psychological. Two views of the same thing reinforce one another. Most of us are able to remember the multitudinous formulas of analysis mainly because we attach to each a geometric picture that keeps us from going astray. Even better than that, the geometric view of a problem helps us to focus on the invariants and to weed out the irrelevant details. For example, a poor choice of a coordinate system may lead to a horrible mess in the analytic formulation, but with some geometric insight, we may be able to choose a much better coordinate system.

Since students are not getting enough geometry in school, one of the avenues left is self-study. Any new attractive books in this field are indeed very welcome. *How does one cut a triangle*? is one of these and is mostly combinatorial geometry that is accessible to bright high school and college students even without too much prior knowledge. Also, there is quite a number of unsolved problems which will challenge professional mathematicians and other scientists. There are various monetary awards for the first received acceptable solutions of some of these latter problems. Like the challenge of climbing mountains because they are, there is a mental challenge of solving the problems not only because they are here (in this book), but also because they are simple to comprehend but still challenging.

Two of the opening problems are "**Grand Problem I. Find all positive integers n, such that every triangle can be cut into n triangles congruent to each other**," and "**Grand Problem II. Find all positive integers n, such that every triangle can be cut into n triangles similar to each other.**" Even deciding which of these two problems is harder is a problem. Although Dieudonné is strongly opposed to "triangles," he may be attracted to the solutions of these problems since one of them requires quite a bit of linear algebra. However, I think Soifer's excursion into the linear algebra necessary in the solution is perhaps too brief.

Another set of problems deals with the problem of given a certain number of points within or on a triangle to determine the maximum of the minimum area formed by three of the points. In particular, three proofs are given to show that for any five points in a triangle of area 1, there are three points that form a triangle of

area not exceeding 1/4. A neater solution appeared in 1980 by Yang and Zhang as a lemma [22]†. With respect to points in a square, see also [23], [24]; with respect to a disk, see "*100 Research Problems in Discrete Geometry*," 1986, #7 (privately printed by W. O. Moser, McGill University. This is an excellent reference source).

The final set of problems and conjectures deal with extremal properties of convex sets. A number of these are due to Paul Erdös. One of the interesting functions defined here is $S_\alpha(F)$. If $0 < \alpha < 1$ and F is a convex figure of unit area, then $S_\alpha(F)$ is the smallest $n \geqslant 3$ such that if n points lie within or on the boundary of F, there are three of these points which form a triangle of area $\leqslant \alpha$. Andy Liu notes that a more informative function would be $\alpha_F(n)$ = the maximum of the minimum area triangle formed from 3 of the n points.

I like the style of the book. In this regard, I agree with Cecil Rousseau who says "it is an adventure story, complete with interesting characters, moments of exhilaration, examples of serendipity, and unanswered questions." Although I gather from this book that Soifer is an enthusiastic teacher, which is a big plus in teaching, I would have preferred him being a little more modest in his writing.

Finally, I would have preferred to see the book more moderately priced since its main target is "young, talented mathematicians who are still in high school or college" and also since it is paperbound of modest length.

REFERENCES

1. New Thinking in School Mathematics, Organization for European Economic Cooperation, 1961.
2. B. Burn, The shape of geometry, *Math. in School*, 19 (1990) 38–39.
3. H. Freudenthal, Mathematics as an Educational Task, Reidel, Dordrecht, 1973, pp. 401–511.
4. S. S. Chern, From triangles to manifolds, *Amer. Math. Monthly*, 86 (1979) 339–349.
5. S. S. Chern, What is geometry?, *Amer. Math. Monthly*, 97 (1990) 679–686.
6. I. M. Yaglom, Elementary geometry, then and now, The Geometric Vein, ed., C. Davis, B. Grünbaum, F. A. Sherk, Springer-Verlag, N.Y., 1981.
7. P. J. Kelly, N. E. Ladd, Geometry, Scott, Foresman, Chicago, 1965 (contains a packet of stereograms for 3-dimensional viewing).
8. V. G. Boltyanskii, Equivalent and Equidecomposable Figures, D. C. Heath, Boston, 1963.
9. E. B. Dynkin and V. A. Uspenskii, Multicolor Problems, D. C. Heath, Boston, 1963.
10. L. I. Golovina and I. M. Yaglom, Induction in Geometry, D. C. Heath, Boston, 1963.
11. B. Yu. Kogan, The Application of Mechanics to Geometry, University of Chicago Press, Chicago, 1974.
12. L. A. Lyusternik, Shortest Paths, Variational Problems, Macmillan, N.Y., 1964.
13. I. F. Sharygin, Problems in Plane Geometry, Mir, U.S.S.R., 1988.
14. I. F. Sharygin, Problems in Solid Geometry, Mir, U.S.S.R., 1986.
15. Yu. A. Shreider, What is Distance?, University of Chicago Press, Chicago, 1974.
16. N. Vasilyev and V. Gutenmacher, Straight Lines and Curves, Mir, U.S.S.R., 1980.
17. I. M. Yaglom, Geometric Transformations, I, II, III, Random House, N.Y., 1962, 1968, 1973.
18. H. Poincaré, The Foundation of Science, Science Press, Lancaster, 1964, p. 431.
19. R. G. Douglas, ed., Toward a Lean and Lively Calculus, M.A.A., Washington, D.C., 1986.
20. L. A. Steen, ed., Calculus for a New Century, A Pump not a Filter, M.A.A., Washington, D.C., 1988.
21. N. Steenrod, The geometric content of freshman and sophomore mathematics courses, CUPM Geometry Conference Report, II, pp. 1–52, October 1967.
22. L. Yang and J. Zhang, The problem about six points in a square, Lectures in Mathematics, Vol. II, (Chinese), Sichuan Renmin Publishing House, 1980, pp. 159–175. An English translation appears in Mathematics Olympiads in China, Hunan Education Publishing House, 1990, pp. 37–49.
23. Problem 745, *Math. Mag.*, 43 (1970) 170–171.
24. M. Goldberg, Maximizing the smallest triangle made by points in a square, *Math. Mag.*, 45 (1972) 135–144.

†I am indebted to my colleague Andy Liu for this reference.

Concrete Mathematics: A Foundation for Computer Science. By Ronald Lewis Graham, Donald Ervin Knuth, and Oren Patashnik, Addison-Wesley, 1989, xiii + 625 pp.

EDWARD A. BENDER
Center for Communications Research, San Diego, CA 92121

Most current calculus texts could almost have been created from one another with scissors, tape, and a color Xerox machine. Attempted innovations fall by the wayside. Rather than being due to the perfection of the standard calculus course, this state of affairs seems to me to be a sequence of the post-Sputnik push for more engineers and for improved science education. This led to a huge demand for "service courses" having a fairly standard list of "must-cover" topics. Many of the intended students had little interest in mathematics for its own sake. These evolutionary pressures shaped the standard calculus text into a massive tome, designed to program uninterested students in the symbol-recognition and rote-manipulation skills required by various science courses.

The proliferation of computers has led to a demand for computer scientists dwarfing the post-Sputnik demand for engineers. There is a list of must-cover mathematical topics for computer science students that has been drawn from combinatorics, logic, mathematical linguistics, and probability theory. Many computer science majors are not interested in mathematics per se. The setting is familiar. Is the basic mathematics course for computer scientists doomed to the same fate as the basic calculus course for engineers and scientists? The spate of new books on "discrete mathematics" suggests that it is.

There is another strategy for teaching the mathematics need in computer science: Instead of relying only on high school mathematic training, build on the mathematical maturity attained by surviving a basic college calculus sequence. The mathematics can then be covered in the depth needed to do computer science instead of at the more superficial level needed to understand computer science textbooks. Graham, Knuth, and Patashnik apparently subscribe to this strategy.

The list of must-cover topics is, however, too long for in-depth coverage in a single course. One obvious approach to this problem is to break the courses down by mathematical discipline; another is to give a brief introductory course and then teach the remaining mathematics as needed in the computer science courses.

What approach is taken in the vaguely titled *Concrete Mathematics: A Foundation for Computer Science*? Neither. The text crosses mathematical disciplines, including material from combinatorics, probability theory, and number theory; the material is not integrated with computer science—an extended discussion of hashing is the only direct applications—and the level is deeper than a brief introductory course. Instead, the manipulative skills and associated concepts that are useful in doing computer science provide the unifying theme. The text grew out of a course initiated by Don Knuth in order to teach computer science students "what he wished somebody had taught him" and is primarily a much more leisurely treatment of material in Section 1.2 of the *Art of Computer Programming* [**1**]. A more informative title might be *Concrete Mathematics: The Computational Foundations of Computer Science*.

The approach in *Concrete Mathematics* is successful.

Certain otherwise strange features of the text can be understood in light of the computational approach. To mention a few: The emphasis on computational skills as well as concepts leads to rather long chapters. Although some enumerative combinatorics is discussed, graph theory is not, presumably because its concepts are not inherently computational but instead tend to rely on other areas of mathematics for computational tools. On the other hand, a chapter is devoted to number theory (which plays less of a role in computer science than does graph theory) because some computational skills in number theory are important in doing computer science.

I was saddened to find that the combinatorial understanding of generating functions is left at the Pólya picture-writing stage—one may want to use Wilf's *Generatingfunctionology* [2] as a supplementary text. Unfortunately, much of the ninety-page chapter that is devoted to manipulating binomial coefficients will need revision in light of the recent work of Zeilberger and Wilf. I was surprised to learn in Section 8.2 that Chebyshev's inequality is the main reason for computing variance—I think that the heuristic result given by assuming asymptotic normality is as important as the hard bounds given by Chebyshev. In contrast to these quibbles, there are many important pleasant surprises in this well-written text. For example:

- Careful attention is given to imparting the ability to usefully manipulate the sorts of expressions that arise in the analysis of algorithms. A student who has to do some manipulations would profit from the relevant chapters. ("Student" is too limited—I learned some interesting things.) Particularly noteworthy are the chapters on sums and integer functions because they contain important ideas and techniques that I haven't seen in any other text.
- There is an introduction to asymptotic analysis which should make students reasonably comfortable with big-oh notation and teach them techniques for estimating simple sums. (Perhaps a future edition will include some discussion of obtaining estimates from generating functions.)
- There is a nice selection of problems with ratings from routine to research level, with solutions and/or literature references included. Many of these are not the boring crank-turning type that flood the standard calculus course.
- A bibliography with over 300 entries is a useful resource for the student or instructor who wishes to pursue a subject further.

Even if you are not interested in considering this book for a course or in reading it yourself, I suggest you at least look at the marginal notes. Many have been written by students, and these range from complaints to witticisms to helpful hints for other students. This idea might liven up some calculus texts more than any number of colored pictures, historical notes or "applications."

In view of the broad mathematical basis of computer science, I hope some of you will regard this text as a challenge to help provide a diversity of well-written texts on the mathematical foundations for doing computer science.

REFERENCES

1. D. E. Knuth, The Art of Computer Programming, volume 1: Fundamental Algorithms, 2nd edition, Addison-Wesley, 1973.
2. H. S. Wilf, Generatingfunctionology, Academic Press, 1990. Reviewed in this MONTHLY, 97 (1990) 864–866.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

Mathematics Appreciation, T(13-14: 1). *The Mathematical Palette.* Ronald Staszkow, Robert Bradshaw. Saunders College, 1991, xiii + 534 pp, $32. [ISBN: 0-03-033274-5] A visually attractive, lively, and appealing introduction to mathematics as a liberal arts subject, emphasizing historical and practical approaches. Assumes some command of symbolic algebra (through elementary trigonometry), but most content is reviewed in the text. Calculator use—together with a short history of mechanical calculation—is reviewed in an appendix. Contents include liberal arts perspectives on algebra, geometry and trigonometry, logic, finance, probability and statistics, and calculus. Basic ideas, rather than detailed calculations, are stressed. Chapters include suggested study projects, with suggested references. PZ

Precalculus, T(13). *Precalculus, Second Edition.* Jerome E. Kaufmann. PWS-Kent, 1991, xv + 640 pp, $35. [ISBN: 0-534-92474-3] Covers traditional topics with emphasis on equations and inequalities, solving problems, graphing techniques, and functions; this edition includes two new sections and minor changes to several others. Supplements include answer key, test banks (printed and computerized), and a TrueBASIC precalculus package. JNC

Precalculus, T(13: 1). *Algebra for College Students, Third Edition.* Bernard Kolman, Arnold Shapiro. Harcourt Brace Jovanovich, 1991, xv + 675 pp, $32. [ISBN: 0-15-502162-1] Retains structure of earlier editions while including new emphasis on calculator usage and mathematical modeling, earlier introduction of exponent notation and linear equations and inequalities, and added sections on linear programming and probability. JNC

Precalculus, T(13: 1). *College Algebra, Second Edition.* Thomas W. Hungerford, Richard Mercer. Saunders College, 1991, xv + 570 pp, $34. [ISBN: 0-03-054243-X] This major rewriting includes earlier coverage of complex numbers, easier method of graphing polynomial functions, more material on conic sections and new sections on variation, partial fractions, matrix algebra, systems of inequalities, linear programming, sequences, infinite series, and probability. A *Student Solutions Manual with Graphing Calculator Supplement* is available. (*First Edition,* TR, August-September 1982.) JNC

Education, T*, S, P. *Geometry for the Classroom.* C. Herbert Clemens, Michael A. Clemens. Springer-Verlag, 1991, xiii + 335 pp, $29.95 (P). [ISBN: 0-387-97564-0]; *Geometry for the Classroom: Exercises and Solutions,* vi + 167 pp, $19.95 (P). [ISBN: 0-387-97565-9] An innovative text in workbook format; presents concepts intuitively, then more rigorously in the construction, proof (stresses rigid motions) and computer program (uses LOGO) sections. Appropriate for high school students, but would also prepare elementary teachers to implement recommended changes in geometry teaching. JNC

Foundations, S, L*. *Philosophy and Mathematics: From Plato to the Present.* Robert J. Baum. Freeman Cooper, 1973, x + 320 pp, $31.25 (P). [ISBN: 0-87735-514-2] Re-issue of a useful anthology of classic sources of philosophical commentary on things mathematical, including relevant excerpts from Plato, Aristotle, Descartes, Hobbes, ..., Kant, Mill, Frege. First published in 1973 (TR, May 1974). LAS

Foundations, T*(13-14: 1), L. *Introduction to Abstract Mathematics, Second Edition.* John F. Lucas. Ardsley House, 1990, x + 382 pp, $41.95. [ISBN: 0-912675-73-X] Well-written text for introduction to proof or discrete mathematics class. Covers logic, proof strategies, sets and cardinality, relations, functions. Some examples and exercises assume knowledge of first-semester calculus. More emphasis on formal logic than most comparable texts. KES

Number Theory, P, L. *Cryptology and Computational Number Theory.* Ed: Carl Pomerance. Proc. of Symposia in Appl. Math., V. 42. AMS, 1990, xi + 171 pp, $57. [ISBN: 0-8218-0155-4] Proceedings of an AMS short course. Contains excellent articles on primality testing (A.K. Lenstra), factoring (Pomerance), discrete logarithms (McCurley), knapsack systems (Odlyzko), provably secure sys-

tems (Goldwasser), and pseudorandom number generators (Lagarias). SG

Linear Algebra, T(14: 1). *Elementary Linear Algebra, Fourth Edition.* Stanley I. Grossman. Saunders College, 1991, xvii + 588 pp, $42.50. [ISBN: 0-03-031193-4] Includes new material on graph theory and least squares applications; additional figures, historical essays, and geometric interpretations; new chapter summaries; introduction of all systems of equations material before matrices; and recurring appearances of a "Summing Up Theorem." The *Applications Supplement* also contains new examples and exercises. (*First Edition,* TR, April 1980; *Second Edition,* TR, January 1985; *Third Edition,* TR, August-September 1987.) JNC

Linear Algebra, T*(13: 1). *Elementary Linear Algebra, Sixth Edition.* Howard Anton. Wiley, 1991, xv + 570 pp, $49.95, [ISBN: 0-471-50900-0]. *Elementary Linear Algebra: Applications Version, Sixth Edition.* Howard Anton, Chris Rorres. Wiley, 1991, xv + 795 pp, $52.95. [ISBN: 0-471-52739-4] Revisions include earlier coverage of eigenvalues and eigenvectors, rewritten material on linear transformations, bases, row space, rank, etc., a separate chapter on inner product spaces, coverage of least squares fitting, new exercises and drill problems. The *Applications Version* (which adds an applications chapter) includes a section on fractals. (*First Edition,* TR, March 1973; Extended Review, 1974; *Fifth Edition,* TR, August-September 1987.) JNC

Algebra, P. *Topics in Computational Algebra.* Eds: G.M. Piacentini Cattaneo, E. Strickland. Kluwer Academic, 1990, 261 pp, $99. [ISBN: 0-7923-1117-5] Papers of ten invited speakers at the Semester of Computational Algebra, University of Rome II, 'Tor Vergata,' January-May 1990. LC

Calculus, T*(13: 2). *Introduction to Mathematical Analysis for Business and Economics.* André L. Yandl. Brooks/Cole, 1991, xxxii + 1000 pp, $52.75. [ISBN: 0-534-13494-7] Also available as two separate volumes: *Finite Mathematics* and *Applied Calculus.* Reviews algebra and functions, then provides two separate paths: systems of linear equations and inequalities, matrices, linear programming, sequences and mathematics of finance, probability; the derivative, applications of derivatives, integration, applications of integration, functions of several variables. Replete with detailed explanations, examples and exercises (many illustrating applications), optional topics. Readable, attractive, pedagogically effective. Worth a look if you have this kind of course. DFA

Algebraic Geometry, P. *Linear Algebraic Groups, Second Enlarged Edition.* Armand Borel. Grad. Texts in Math., V. 126. Springer-Verlag, 1991, xi + 288 pp, $49. [ISBN: 0-387-97370-2] An enlargement of the 1969 Benjamin edition. New sections include discussions of Cartan subgroups of solvable groups, isotropic reductive groups, and Tits systems. SG

Geometry, P. *Integral Geometry and Tomography.* Eds: Eric Grinberg, Eric Todd Quinto. Contemp. Math., V. 113. AMS, 1990, xii + 249 pp, $54 (P). [ISBN: 0-8218-5120-9] Twenty papers from the Joint Summer Research Conference held at Humboldt State University, Arcata, California, June 24-30, 1989. Subjects represented include combinatorics, geometric inequalities, microlocal analysis, group theory, harmonic analysis, Lie group theory, Radon transforms, radiation dose planning, radar, partial differential equations. DFA

Control Theory, T(18: 1), L. *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach.* Tamer Başar, Pierre Bernhard. Systems & Control: Found. & Applic. Ser. Birkhauser Boston, 1991, xii + 224 pp, $49.50. [ISBN: 0-8176-3554-8] The task at hand in this book is to determine a controller design which optimizes the worst case scenario. If we think of "optimizing" a system as minimizing some cost function, we can think of a random disturbance as a factor seeking to maximize the same function. Thus, a minimax scenario arises, leading to the application of game theory. Written for second level graduate students, this book considers both continuous and discrete time systems. A course in linear control theory is an adequate preparation for this book. SM

Control Theory, T(17: 1), P, L. *Lecture Notes in Control and Information Sciences-152: Implicit Linear Systems.* J.D. Aplevich. Springer-Verlag, 1991, xi + 176 pp, $31 (P). [ISBN: 0-387-53537-3] An introduction to a class of models of linear dynamical systems. These systems are characterized by a single vector of external variables rather than the usual vectors for inputs and outputs. Meant for graduate students in engineering or applied mathematics with some prior experience with control theory. SM

Statistics, P. *Applied Computational Statistics in Longitudinal Research.* Michael J. Rovine, Alexander von Eye. Academic Pr, 1991, xiii + 237 pp, $34.50 (P). [ISBN: 0-12-599450-8] Companion volume to the 1990 two-volume *Statistical Methods in Longitudinal Research,* edited by von Eye (TR, March 1991). Presents fourteen of the newer techniques—summarizes the method, outlines the computational steps, and gives an annotated data example. RSK

Statistics, P.** *Directions in Robust Statistics and Diagnostics.* Eds: Werner Stahel, Sanford Weisberg. Springer-Verlag, 1991. *Part I,* IMA, V. 33, xviii + 252 pp, $35, [ISBN: 0-387-97530-6]; *Part II,* IMA, V. 34, xviii + 380 pp, $45. [ISBN: 0-387-97531-4] Proceedings of the first four weeks of a six-week program of the IMA (Institute for Mathematics and Its Applications) at the University of Minnesota in the summer of 1989. Contains 39 papers by leaders in these areas, giving an overview of research being done. (Parts are divided alphabetically by principal author.) RSK

Statistics, T(17-18: 1, 2), S, P, L. *A User's Guide to Principal Components.* J. Edward Jackson. Wiley, 1991, xvii + 569 pp, $69.95. [ISBN:

0-471-62267-2] Designed for practitioners of principal component analysis (PCA), this text carefully presents the methods of PCA and illustates them with several data sets. Topics include multidimensional scaling, principal components regression and analysis of variance, and applications of PCA to discriminant and cluster analysis, and time series. PCA is contrasted with factor analysis and competitors to PCA are discussed. RWJ

Statistics, P. *Some Recent Results on Chi-Squared Tests.* M.S. Nikulin. Papers in Pure & Appl. Math., No. 86. Queen's Univ, 1991, 74 pp, (P). Divided into three parts: the chi-squared test of Pearson, a chi-squared goodness-of-fit test for exponential distributions of rank one, and chi-squared test for multivariate discrete distributions. RSK

Programming, S(13), P. *Handbook of QuickBASIC.* David I. Schneider. Brady, 1991, 797 pp, $29.95 (P). [ISBN: 0-13-381880-2] A reference manual suitable for beginning and experienced programmers. Covers QuickBASIC, QBASIC, and the BASIC Professional Development System (PDS). Assumes reader knows rudiments of QuickBASIC. Extension Comments and over 700 programming examples accompany the descriptions of functions, statements, and metacomments, as do hundreds of programming tips and techniques. Many useful appendices and tables address special topics. DFA

Programming, T(13: 1), C. *Common-Sense BASIC: Structured Programming with Microsoft QuickBASIC.* Alice M. Dean, Gove W. Effinger. Harcourt Brace Jovanovich, 1991, xvi + 415 pp, $28 (P) net. [ISBN: 0-15-512297-5] A leisurely introduction to rudiments of structured programming in Basic (specific references are to Microsoft QuickBASIC). Assumes no computer background whatever—initial sections include photos of computers, definitions of computer jargon, etc. Emphasizes programming *structure*: control structures, data types, data structures, etc. Final chapters introduce graphics, color, and sound. Chapters end with programming exercises. PZ

Theory of Computation, T(17-18), P, L. *Algebraic and Structural Automata Theory.* Ed: Boleslaw Mikolajczak. Annals of Disc. Math., V. 44. North-Holland (US Distr: Elsevier Science), 1991, xxi + 402 pp, $131.50 (P). [ISBN: 0-444-87458-5] English translation of the 1985 Polish edition. Principal results of the Group in Automata Theory and Computability at the Technical University of Poznan. Addresses problems in automata morphisms, time-varying automata and their transformations, automata realizations, linear automata, input subautomata, cellular networks. Assumes only basic concepts of algebra and graph theory. Exercises, examples, writing style make it quite suitable as a text. But note price. DFA

Applications (Physical Science), P. *Applications of the Mathematical Sciences to Materials Science.* National Research Council. National Academy Pr, 1991, ix + 26 pp, (P). A non-technical review of mathematical challenges and opportunities in the rapidly growing field of materials science, where computation-intensive mathematical modelling is becoming more effective than laboratory trial-and-error. Mathematical methods range broadly from partial differential equations and nonlinear optimization to tiling and cellular automata; applications are equally diverse, including ceramics, polymers, and protein folding. LAS

Applications (Physical Science), T(17-18: 1), S, P, L. *Radar and Sonar, Part I.* Richard E. Blahut, Willard Miller, Jr., Calvin H. Wilcox. IMA, V. 32. Springer-Verlag, 1991, xi + 260 pp, $34. [ISBN: 0-387-97516-0] Thorough, carefully prepared lecture notes from three tutorials on remote surveillance algorithms, applications of harmonic analysis to radar and sonar, and echo structure, plus a previously unpublished (1960) technical report on the synthesis problem for radar ambiguity. Can be used as a textbook. BC

Applications (Physics), P. *Soliton Phenomenology.* Vladimir G. Makhankov. Math. & Its Applic., V. 33. Kluwer Academic, 1990, xi + 452 pp, $149. [ISBN: 90-277-2830-5] Presents the essentials of soliton theory in six parts: Part I—Quantum systems and classical behavior; Part II—Some exact results in one-dimensional space; Part III—Noncompact symmetries and Bose gas; Part IV—Soliton-like solutions in one-dimension; Part V—Phenomenology of $D = 1$ solitons; and Part VI—Many-dimensional solitons. MPR

Applications (Physics), P. *The Geometry and Physics of Knots.* Michael Atiyah. Cambridge Univ Pr, 1990, x + 78 pp, $39.50. [ISBN: 0-521-39521-6] Notes expanding on a series of lectures delivered at the University of Florence, November 1988, offering a mathematical introduction to recent applications of knot theory in theoretical physics. The Jones polynomial invariant of a knot finds clear conceptual interpretation in Witten's topological quantum field theory. RB

Applications (Physics), S(16-18), P. *Wave Packets and Their Bifurcations in Geophysical Fluid Dynamics.* Huijun Yang. Appl. Math. Sci., V. 85. Springer-Verlag, 1991, ix + 247 pp, $39. [ISBN: 0-387-97257-9] This monograph is conspicuously well-written. Mathematical formulae are embedded in text that is both easy to read and highly illuminating. The subject matter is introduced carefully, making it accessible to the non-expert. MU

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; SG: Steven Galovich, Carleton; RWJ: Roger W. Johnson, Carleton; RSK: Richard S. Kleber, St. Olaf; SM: Steve McKelvey, St. Olaf; MPR: Matthew P. Richey, St. Olaf; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

Push the power of Mathematica to its limits

Mathematica® in Action

Stan Wagon, *Macalester College*

Mathematica in Action extends your imagination and the power of *Mathematica* by providing alternative methods to generate three-dimensional graphics, iterative graphics, and animations. Its many valuable shortcuts, complete programs with line-by-line explanations, and *hundreds* of advanced examples worked in detail help you realize those aims.

Whether you are a mathematics teacher, researcher, or enthusiast, you'll find this to be *the* indispensable sourcebook to own because of its • detailed examples that use animations to study dynamical systems and Julia Sets • attention to high-precision number theory—including many aspects of prime numbers • and its in-depth study of cycloids.

Contents • A Brief Introduction • Prime Numbers • Rolling Circles • Surfaces • Iterative Graphics • Iterative Complex Graphics • The Turtle Road to Recursion • Advanced Three-Dimensional Graphics • Some Algorithms of Number Theory • Imaginary Primes and Prime Imaginaries • Additional Examples • Appendix: Supplementary Programs

1991, 419 pages, 133 illustrations, 4 pp. color plates; paper (2202-X) $29.95; cloth (2229-1) $39.95

A remarkable visual exploration of the world of fractals and chaos in a 63-minute, full-color video—

FRACTALS

An Animated Discussion with EDWARD LORENZ and BENOIT B. MANDELBROT

H.-O. Peitgen • H. Jurgens • D. Saupe • C. Zahlten

The video turns the Mandelbrot set and the Lorenz attractor into visible and easily comprehensible objects as their discoverers, Benoit B. Mandelbrot and Edward Lorenz, discuss the background, history, and details of their work in extended interviews.

1991, VHS, color video, 63 minutes, ISBN 2213-5, $59.95 individual price; $149.95 educational price (includes licensing and duplication rights)

Mail to: **W. H. Freeman and Company**

4419 West 1980 South, Salt Lake City, UT 84104

Send me____copies of **Mathematica in Action,**

paper (2202-X), at $29.95 _____

hardbound (2229-1), at $39.95 _____

Send me____copies of **FRACTALS,** (2213-5), individual price at $59.95 _____

Educational price at $149.95 (includes licensing and duplication rights) _____

Add $1.95 shipping/handling for the first item _____

$1.25 for each additional item _____

UT, CA, & NY residents, please add the appropriate sales tax _____

Total _____

☐ I enclose a check or money order payable to **W. H. Freeman and Company**

☐Charge to my ☐MasterCard ☐Visa Exp. date _____

Account #_____

Signature _____

(credit card orders must be signed)

Name _____

Address _____

City_____

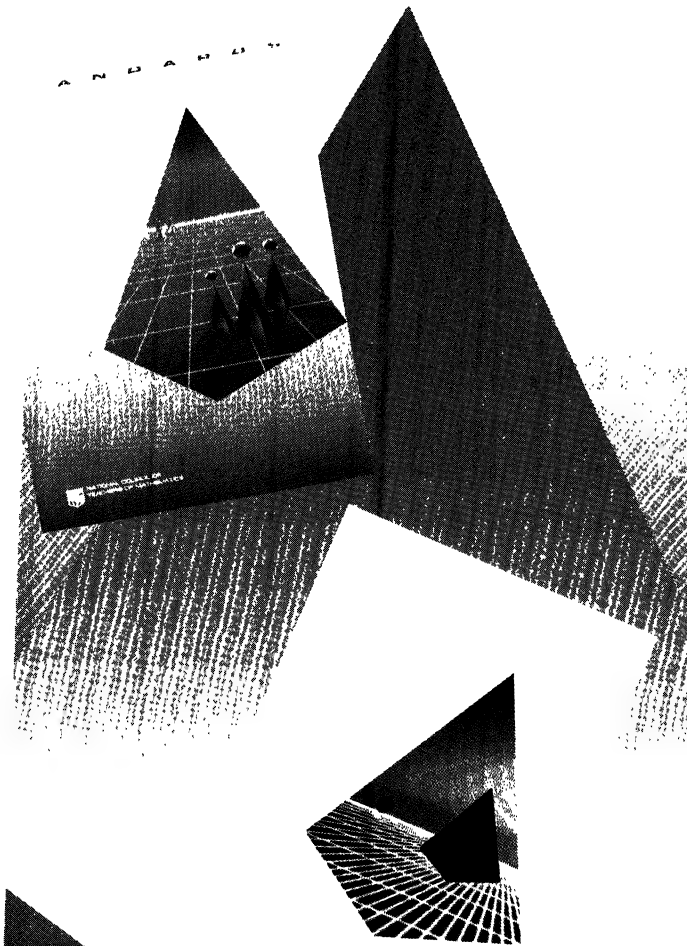
State/Zip _____

J712

W. H. Freeman and Company

The book publishing arm of Scientific American

Be a "Standards-bearer" and spread the news!



You now have available to you the two important resources for improving the teaching and learning of mathematics in the 1990s.

This two-document set will help you chart programs that will increase the mathematical power of all students—from kindergarten through high school.

Special price for the two-book set, **$42.50**
ISBN 0-87353-308-9, #481S6

Professional Standards for Teaching Mathematics
200 pp., ISBN 0-87353-307-0, #439S6, $25*

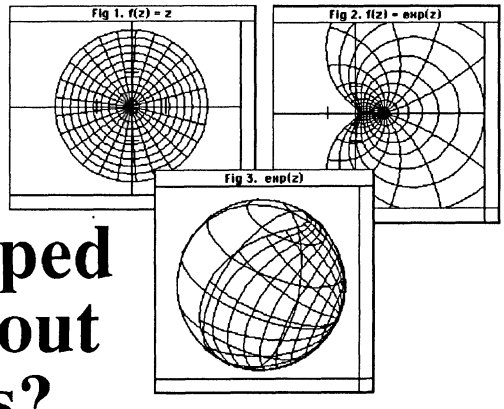
Curriculum and Evaluation Standards for School Mathematics
258 pp., ISBN 0-87353-273-2, #396S6, $25*

Quantity discounts for each book are also available.

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS
1906 Association Drive, Reston, VA 22091
Tel. (703) 620-9840 or fax (703) 476-2970
For orders only, call (800) 235-7566

*$20 for NCTM individual members.

When was the last time a computer program helped you think about mathematics?



Software and video tapes for the student, professional, and anyone who loves mathematics. ***Ask for our latest catalog.***

Lascaux Graphics (602) 544-4229 **(800) 338-0993**
7601 N. Calle Sin Envidia, Suite 31 - Tucson, AZ 85718 USA

Introducing E.Z. Math, E.Z. Algebra and E.Z. Arithmetic for the HP 48SX

E.Z. Math, E.Z. Algebra and E.Z. Arithmetic are programs for the Hewlett Packard 48SX calculator conceived, written and programmed by Raymond La Barbera and the E.Z. Software Company. Each program comes on a 128K plug-in ROM card accompanied by an easy-to-understand, well-written, detailed manual loaded with lots of specific examples and is designed for use by students, teachers, parents and business people. Each program features an easy-to-use, logically organized, user-friendly interface which enables those who consider themselves to be calculator and computer illiterates, as well as those who don't like to read manuals, to have full access to all program features quickly and easily. Since the HP 48SX is essentially an impressive, sharp-looking, 8 ounce pocket computer, students are easily motivated to take it along with them to study, practice, drill and master math in a study hall, on a train or bus, in a car, on line, on vacation, on a break—in short, for self-study at any time and in any place.

What Can Be Done With E.Z. Math

E.Z. Math effectively solves problems involving graphs, numbers, loans and savings. With E.Z. Math, anyone can:
- Master the entire high school and college graphing curriculum, from algebra to calculus, with 188 families of equations, inequalities, functions, and systems, all arranged in an easy-to-use, user-friendly system of menus to make graphic analysis a snap!
- Get extensive help with calculations involving fractions, whole numbers, complex numbers and number sequences.
- Easily do savings and loan calculations and generate complete amortization tables.
- Learn many basic concepts including those involving sets, variables, graphing, solving, numbers, loans and savings.

What Can Be Done With E.Z. Algebra

E.Z. Algebra is a comprehensive ninth grade high school basic algebra course as well as a high school and college remedial algebra course that builds a solid algebra foundation. With E.Z. Algebra, anyone can:
- Learn about sets, operations, variables, relations and other concepts essential to a real understanding of algebra.
- Understand the sets of natural numbers, whole numbers, integers, rational numbers and real numbers.
- Master the meaning and properties of the operations of addition, subtraction, multiplication, division, power and root.
- Do all kinds of problems involving algebraic expressions, numerical phrases, equations and inequalities.

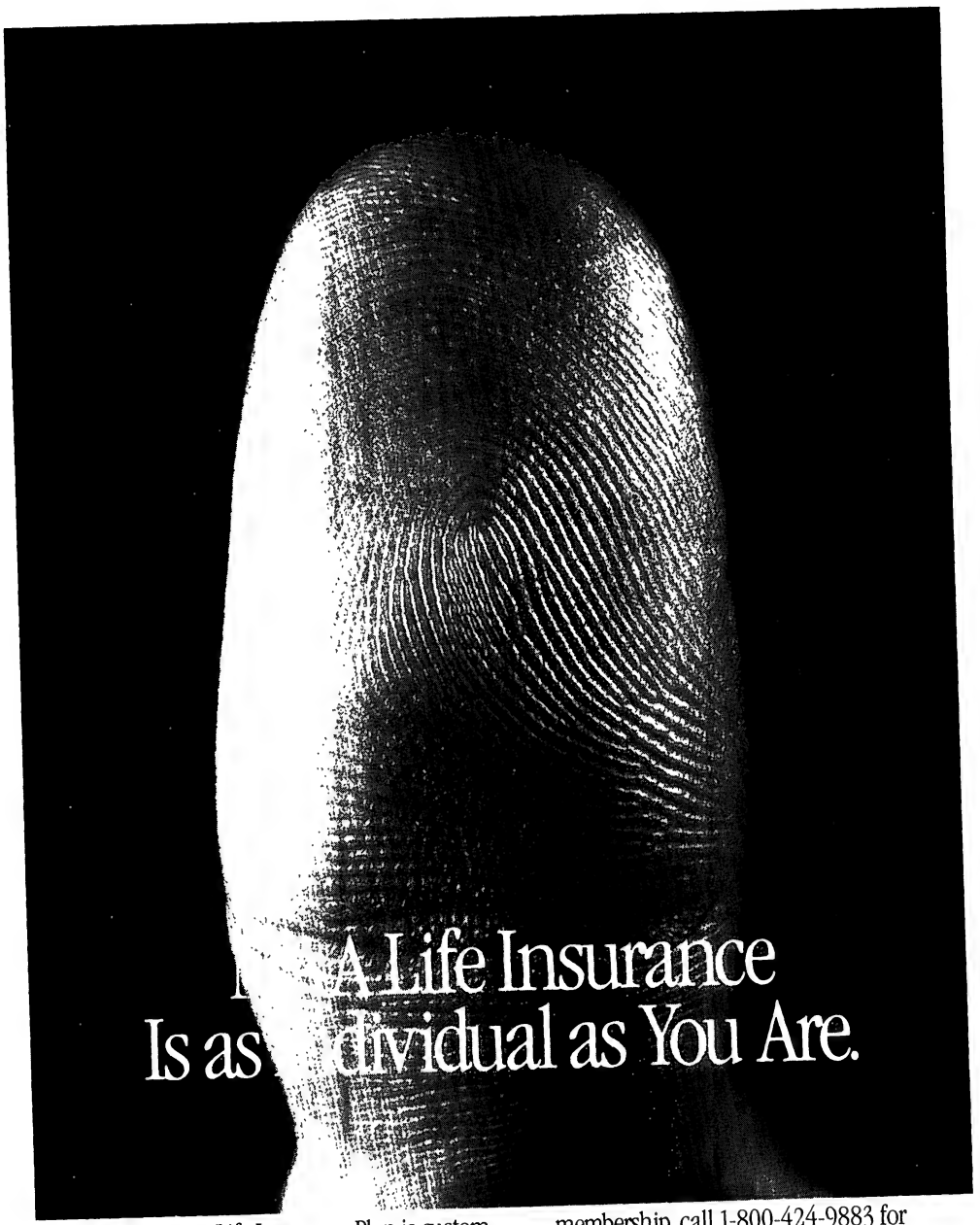
What Can Be Done With E.Z. Arithmetic

E.Z. Arithmetic is a comprehensive elementary school basic arithmetic course as well as a high school and college remedial arithmetic course that makes solving most arithmetic problems a snap! With E.Z. Arithmetic, anyone can:
- Learn how to add, subtract, multiply, divide and order whole numbers, fractions, decimals, percents and integers.
- Master the meaning, terminology and conversion methods for whole numbers, fractions, decimals and percents.
- Drill and be graded on endlessly varied, randomly selected sets of problems involving whole numbers, fractions, decimals and integers, with the difficulty level, number of problems, operation and type of number user selectable.

How To Order Copies or Get Further Information

Each E.Z. Software program costs $130.00 ($125.00 retail, plus $5.00 shipping and handling). Take a 10% discount when ordering ten or more units. We accept payment by check, money order, COD, VISA, MC, AE and purchase order. If within 30 days you find that any E.Z. Software program fails to meet your expectations, we'll gladly take back your copy for a prompt, courteous refund. To order copies, either individually or bundled with HP 48SX calculators, please contact:

SMI Corporation, 250 West New Street, Dept MM2, Kingsport, Tennessee 37660
(800) 234-0123 or (615) 378-4821 or (615) 245-8982 (Fax).



Our Term Life Insurance Plan is custom-designed for members of our profession. What's more, each policy can then be tailored to suit your individual needs. As these needs change, so can the policy—and it can stay with you no matter how often you change jobs.

Our group purchasing power helped us to negotiate top quality insurance, at a very low price. To take advantage of this benefit of membership, call 1-800-424-9883 for further details (in Washington, D.C. call 457-6820).

MAA INSURANCE
Designed by Members.
For Members.

The MAA Life Plan is underwritten by Connecticut General Life Insurance Company, A CIGNA Company, Hartford, Connecticut 06152
on form number GM 3000

ZENITH DATA SYSTEMS PRESENTS
MASTERS OF INNOVATION

Solving The Problem Of How Students Solve Problems.



Dr. Ronald H. Stevens

Any professor will tell you that good teaching requires an understanding of their students' thinking. At the University of California Los Angeles, however, Dr. Ronald H. Stevens has found an especially innovative way to get inside the cranium of his second-year medical students.

With his award-winning "IMMEX" software program, Dr. Stevens discovers not only *if* his students can solve Immunology problems, but also *how* information was gathered and processed during the solving.

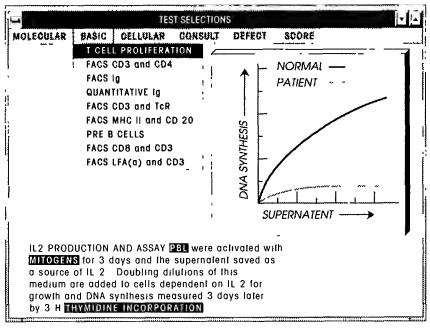
Here's how it works. Programmed in Microsoft® *Windows™ version 3.0,* the easy-to-use "IMMEX" consists of multiple cases of immune defects and a set of results from 45 laboratory tests (see figure below). Through a series of exercises and exams, students are asked to diagnose these cases by selecting the appropriate tests and examining their results.

Upon completion, graphical representations are generated by computer to show which tests were chosen and importantly, demonstrate how students searched for the solution to each problem.

It is then possible to visualize the students' thought process in a way that standard, multiple choice testing doesn't allow. For instance, Dr. Stevens can learn how organized and focused their knowledge is, how well their organization relates to critical concepts in Immunology, where major misconceptions exist and whether proper knowledge links are evident.

In turn, these insights gained through "IMMEX" can have a significant impact on teaching methods. As Dr. Stevens explains, "This approach can lead to rapid detection and remediation of individual students' problem solving difficulties, and can greatly personalize the education process."

What was Dr. Stevens' approach in developing his award-winning software? He chose to program "IMMEX" on Zenith Data Systems laptop PCs. They provided him with all the Random Access Memory and portability required to work after hours and over weekends. And that made solving the problem of how students solve problems, less of a problem.



"IMMEX" Software Program

ABOUT THE MASTERS OF INNOVATION COMPETITION.

As a corporation committed to education, Zenith Data Systems encourages students and educators—like Dr. Ronald H. Stevens —to creatively explore the potential of computers within their fields of study. Towards that end, Zenith Data Systems has sponsored the MASTERS OF INNOVATION Competition for the past three years.

To obtain an unabridged copy of his discussion paper on "IMMEX," or an application to enter the MASTERS OF INNOVATION IV Competition, please write to us at: Masters Of Innovation Program, Zenith Data Systems Corporation, P.O. Box 14513, Chicago, IL 60614-9998.

Microsoft and *Windows* are trademarks of Microsoft Corporation. Copyright © 1991 Dr. Ronald H. Stevens.
Copyright ©1991 Zenith Data Systems Corporation.

Springer for Mathematics

C.H. Clemens, University of Utah, Salt Lake City, Utah and **M.A. Clemens**

Geometry for the Classroom

Intended for use in college courses for prospective or in-service secondary school teachers of geometry. Designed to give teachers broad preparation in the content of elementary geometry as well as closely related topics of a slightly more advanced nature. The presentation and the modular format are designed to incorporate a flexible methodology for the teaching of geometry, one that can be adapted to different classroom settings. The basic strategy is to develop the few fundamental concepts of elementary geometry, first in intuitive form, and then more rigorously. The rest of the material is then built up out of these concepts through a combination of exposition and "guided discovery" in the problem sections. A separate volume including the solutions to the exercises is also available.

1991/335 pp., 512 illus./Softcover/$29.95
ISBN 0-387-97564-0

Geometry for the Classroom; Exercises and Solutions

1991/167 pp./Softcover/$19.95
ISBN 0-387-97565-9

P. Ribenboim, Queen's University, Kingston, Ontario, Canada

The Little Book of Big Primes

This abridged version of **The Book of Prime Number Records,** also by Ribenboim, presents records concerning prime numbers. It also explores the interface between computations and the theory of prime numbers. Contains an up-to-date historical presentation of the main problems pertaining to prime numbers, as well as many fascinating topics, including primality testing. Written in a light and humorous language, this book is thoroughly accessible to everyone.

1991/app. 304 pp./Softcover/$29.50
ISBN 0-387-97508-X

D. Bressoud, Pennsylvania State University, University Park, PA

Second Year Calculus

From Celestial Mechanics to Special Relativity

Covers multi-variable and vector calculus, emphasizing the historical physical problems which gave rise to the concepts of calculus. The book carries us from the birth of the mechanized view of the world in Isaac Newton's *Mathematical Principles of Natural Philosophy* in which mathematics becomes the ultimate tool for modelling physical reality, to the dawn of a radically new and often counter-intuitive age in Albert Einstein's *Special Theory of Relativity* in which the mathematical model suggests new aspects of that reality. The development of this process is discussed from the modern viewpoint of differential forms. Using this concept, the student learns to compute orbits and rocket trajectories, model flows and force fields, and derive the laws of electricity and magnetism. These exercises and observations of mathematical symmetry enable the student to better understand the interaction of physics and mathematics.

1991/app. 416 pp., 98 illus./Softcover/$29.95
ISBN 0-387-97606-X
Undergraduate Texts in Mathematics
Readings in Mathematics

Order Today!

- Call: Toll-Free 1-800-SPRINGER(R): 1-800-777-4643. In NJ call 201-348-4033 (8:30 AM - 4:30 PM EST). Your reference number is S902.
- Write: Send payment plus $2.50 for postage and handling to: Springer-Verlag New York, Inc., Order fulfillment-S902, PO Box 2485, Secaucus, NJ 07096-2491.
- Visit: Your local technical bookstore.

Instructors: Call or Write for information on textbook examination copies!



Springer-Verlag

New York • Berlin • Heidelberg • Vienna • London • Paris • Tokyo • Hong Kong • Barcelona • Budapest

DO MORE WITH THE NEXT GENERATION

After inventing the graphic calculator and watching it enhance the learning process, Casio has gone a step further. We've created The Next Generation.

When students use the fx-7700G in tandem with a teacher's OH-7700G, their minds are focused on the logic and purpose behind calculations, not just the mechanics of doing them.

And Casio didn't just top Casio, we topped our competition. With exclusives like integration with shading. Fractional capabilities. Inequality graphs. And memory protection when batteries are replaced.

INNOVATIONS '91 DESIGN & ENGINEERING AWARD CONSUMER ELECTRONICS SHOW



Plus, compared to our leading competitor, we have almost twice the memory. More zoom modes. 4-way scrolling (up and down, as well as left and right). A menu-key approach that reduces keyboard clutter while increasing functionality. And a powerful, "spread-sheet-like" 9X9 matrix that's easy to view and edit.

The OH-7700G.

Casio's Next Generation. A sure way to assist the teaching and improve the learning of America's next generation.

CASIO

Where Miracles Never Cease®

Casio Educational Product Distributors

ABC SCHOOL SUPPLY
800-669-4ABC
(IN GA 404-497-0001)

ALLIED NATIONAL
800-999-8099
(IN MI 313-543-1232)

ARROWHEAD BUSINESS MACHINES
800-234-3396
(IN TX 214-869-0721)
(IN MO 816-861-1113)
(IN CA 213-946-6680)

THE BACH COMPANY
800-248-2224
(IN CA 415-424-0800)

BARDEEN SCHOOL SUPPLY
800-831-0113
(IN NY 315-437-7566)

BECKLEY-CARDY COMPANY
800-446-1477

BHARDS PUBLISHING
800-473-7999
(IN IL 312-642-8657)

CALCULATORS INC.
800-533-9921
(IN MN 800-533-9921)

CAROLINA WHOLESALE
800-521-4600
(IN NC 704-598-8101)

COLBORN SCHOOL SUPPLY INC.
800-548-7031
(IN CO 303-778-1220)
(IN MT 406-245-3158)

COPCO
800-446-7021
(IN OH 800-589-3006)

CUISENAIRE
800-237-3142
(IN NY 914-235-0900)

DALE SEYMOUR PUBLICATIONS
800-872-1100
(IN CA 800-872-1100)

THE DOUGLAS STEWART CO.
800-279-2795
(IN WI 800-279-2795)

E.A.I.
800-272-0272
(IN NJ 201-891-9466)

EDUCATIONAL ELECTRONICS
800-526-9060
(IN MA 617-821-6458)

GLOBAL PRODUCTS
800-633-0633
(IN IL 708-397-4944)

GREYSTONE EDUCATIONAL MATERIALS
800-733-0671
(IN MN 612-430-9857)

KURTZ BROTHERS
800-252-3811
(IN PA 814-765-6561)

LONGINO DISTRIBUTORS
800-633-6224
(IN NC 704-873-3282)

NASCO
800-558-9595
(IN WI 414-563-2446)

NATIONAL AUDIO-VISUAL SUPPLY
800-222-0109
(IN NJ 800-222-0109)

NEMESIS DISTRIBUTING
800-940-7407
(IN FL 305-477-8822)

PENNS VALLEY PUBLISHING/ LEARNING SYSTEMS
800-422-4412
(IN PA 215-855-4948)

SERVCO PACIFIC
(IN HI 808-841-7566)

TAM'S
800-421-5188
(IN CA 800-244-5624)

TAYLOR ELECTRIC
800-558-6970
(IN WI 800-242-8940)

TECHLINE
800-777-3635
(IN VA 703-389-0857)

TROXELL COMMUNICATIONS INC.
800-528-7912
(IN AZ 800-352-7941)

VALLEY BUSINESS MACHINES
(IN UT 801-969-6303)

VISTATECH
800-847-9851
(IN NY 212-254-9851)
(IN CA 213-602-0277)

WHOLESALE ELECTRONIC SUPPLY
800-527-2156
(IN TX 800-441-0145)

CASIO.

THE MATHEMATICAL SYMBOL FOR QUALITY

NEW FOR 1992

Calculus and Analytic Geometry, 5/e
Sherman Stein & Anthony Barcellos

Calculus for Business, Economics, and the Social and Life Sciences, 5/e
Laurence Hoffmann & Gerald Bradley

Calculus Gems: Brief Lives and Memorable Mathematics
George Simmons

Discovering Calculus with the HP-28 and the HP-48
Robert Smith & Roland Minton

Introduction to MathCAD for Scientists and Engineers
Sol Wieder

Encounters with Chaos
Denny Gulick

Foundations of Abstract Mathematics
David Kurtz

Numerical Methods and Analysis
James Buchanan & Peter Turner

Statistical Methods in the Biological and Health Sciences, 2/e
J. Susan Milton

Fundamentals of Computing: Volume 1 Logic, Programs, and Computing
Allen Tucker, James Bradley, Robert Cupper, & David Garnick

Also of Interest from McGraw-Hill

Applications of Discrete Mathematics
Michaels/Rosen

Discrete Mathematics and Its Applications, 2/e
Rosen

Differential Equations with Applications and Historical Notes, 2/e
Simmons

Partial Differential Equations and Boundary Value Problems with Applications, 2/e
Pinsky

For more information, please contact your McGraw-Hill representative or write to McGraw-Hill College Division, Comp Processing & Control, P.O. Box 448, Hightstown, NJ 08520.



Wadsworth & Brooks/Cole Advanced Books & Software

State-of-the-art software

MathWriter™ 2.0: The Scientific Word Processor for the Macintosh®
J. Robert Cooke and E. Ted Sobel

MathWriter is a complete "what-you-see-is-what-you-get" scientific word processing program that allows users to enter mathematical expressions as text, not graphics, so both text and mathematics can be edited in the same document. In addition to a spelling checker, thesaurus, and automatic hyphenation, the program also automatically sizes and centers mathematical symbols, italicizes variables, and formats tables and matrices. **MathWriter** shows onscreen renumbering of both equations and references to equations and can show both old and revised versions of the same document.

1991. Professional version: $395. Reduced-featured Educational Version $99.95.

EXP: The Scientific Word Processor, Version 2.1
Simon L. Smith

This easy-to-use, "what-you-see-is-what-you-get" technical word processing program for IBM PCs and compatibles offers fast, professional results. Used by mathematicians and scientists worldwide, EXP has the built-in capabilities of a complete word processor, near typeset quality output, the ability to import graphics and see them on screen, and much, much more! The new ***EXP DVI Output Driver*** allows an EXP file to be saved as a "device independent" file and printed on any device with TeX fonts available. Now EXP files can be printed at typeset resolutions up to 2,000 dots per inch.

August 1991. Single User Package $295. Upgrade package $35 for registered users of Version 2.0 and $150 for registered users of any Version 1.0. DVI Output Drive Version 2.1 $195.

Advanced Books

Fourier Analysis and Its Applications
Gerald B. Folland

This undergraduate text presents Fourier analysis as a powerful tool for solving various differential equations of interest in science and engineering. The comprehensive work covers the whole realm of Fourier analysis including Fourier series and transforms, eigenfunction expansions, Laplace transforms, and Green's functions.

December 1991. ISBN: 0-534-17094-3. $59.50.

The Geometry of Computer Graphics
Walter Taylor

This senior-level text presents the power and beauty of pure geometry in a way that it can be applied. The text discusses how one commands a computer to make pictures. It presents the fundamental mathematical concepts needed for two- and three-dimensional computer graphics, the various languages in which these ideas are communicated, and many ways to describe pictures mathematically. Included are a wealth of exercises, from elementary problems to major works in analytic and projective geometry that will both test and challenge students.

October 1991. ISBN: 0-534-17100-1. $58.95.

A Course in Ring Theory
Donald S. Passman

This graduate textbook offers a module theoretic stroll through a mixture of commutative and non-commutative ring theory, with emphasis on the latter. Projective and injective modules supply the framework for a coherent choice of material.

1991. ISBN: 0-534-13776-8. $54.95.

Ethnomathematics: A Multicultural View of Mathematical Ideas
Marcia Ascher

This unique book discusses the mathematical ideas of people in traditional, or "small-scale," cultures such as American Indian tribes, Incas, and various cultures in the South Pacific and Africa. The book takes some common concepts such as number, chance, and geometry, and describes the role of these ideas in the cultures.

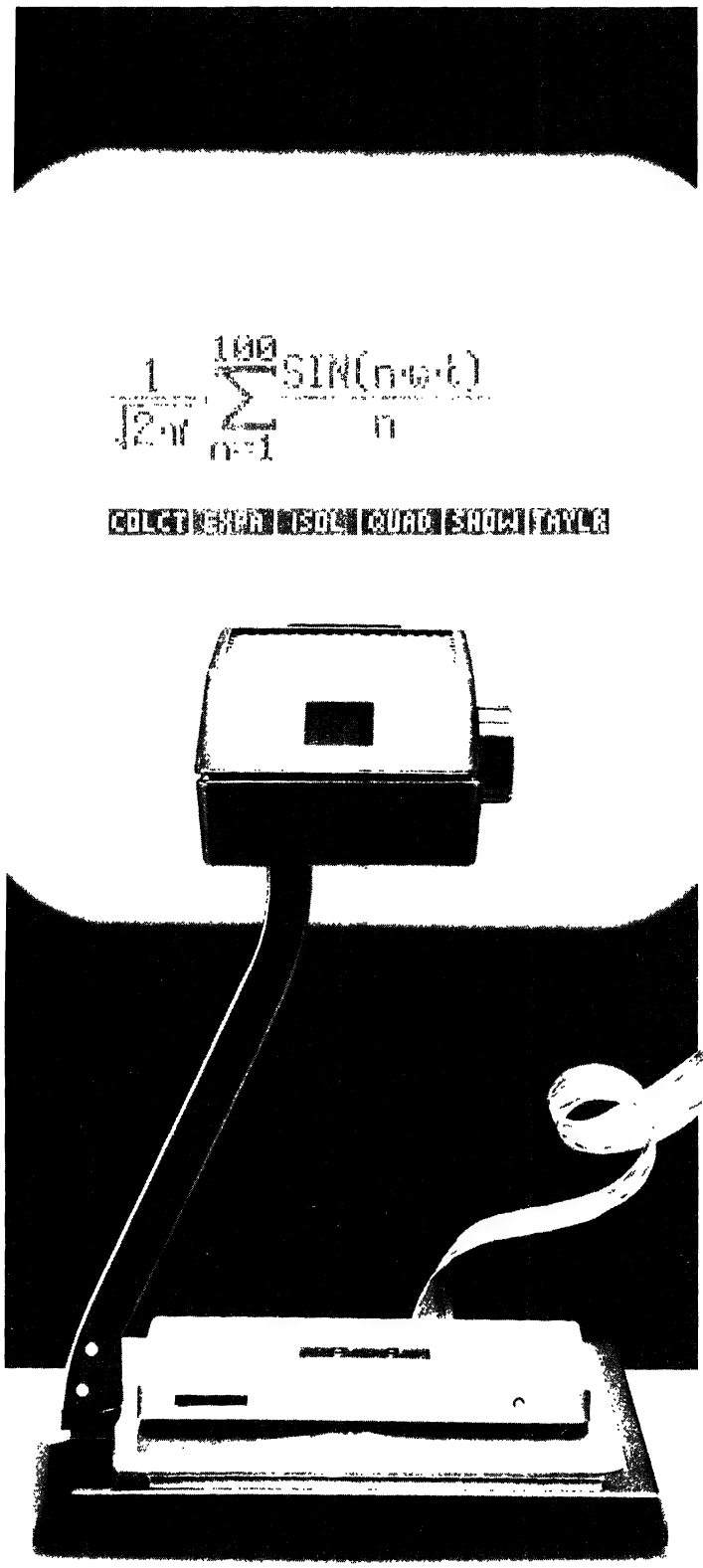
1991. ISBN: 0-534-14880-8. $38.95.

To order a personal copy of software or books, use our toll-free number (800) 354-9706, or write to us at the address below. To request complimentary copies for review, or to receive a copy of our 1992 Advanced Books & Software Catalog, please write:

Wadsworth & Brooks/Cole Advanced Books & Software

MM002
511 Forest Lodge Road
Pacific Grove, CA 93950-5098

Help your students discover more meaningful relationships.



Renewed for '91: a free classroom display device with purchase of 30 calculators.

Showing is much more powerful than telling. So we've developed special classroom displays for our most advanced calculators.

The HP 48SX scientific expandable calculator, the cost-effective HP 28S and the new HP 48S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculators.

If your department or students purchase 30 HP 48SX, HP 28S or HP 48S calculators (or a mix of all three), we'll give you free an HP 48SX and plug-in classroom display (a $900 retail value). Or a classroom display version of the HP 28S (a $600 retail value). And free teaching materials.

Call **(503) 757-2004** from 8am to 3pm PDT for details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1991, and applies only to college and high school instructors.

HEWLETT PACKARD

NEW! VERSION 2

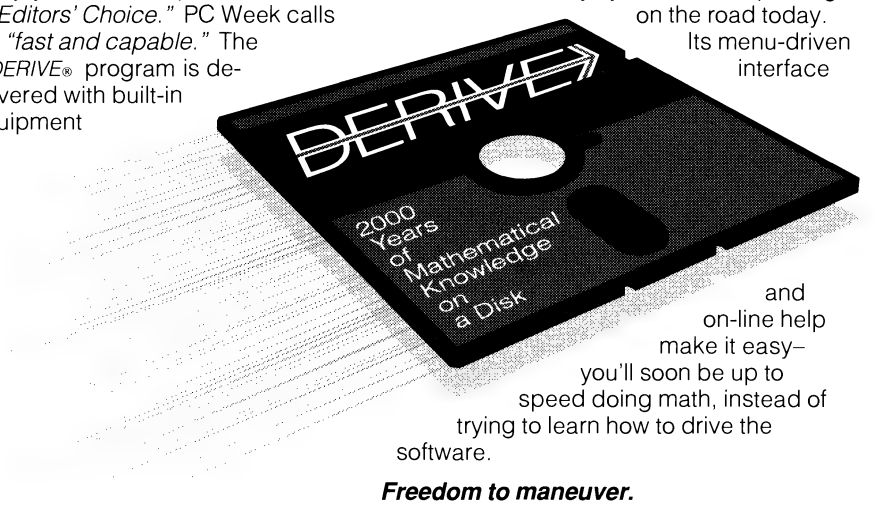
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE® ing.

PC MAGAZINE EDITORS' CHOICE
May 29, 1990
Derive Version 1.6

DERIVE® – A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy– you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 9 November 1991

Contents

(ISSN 0002-9890)

ARTICLES

Box-Spline Tilings CARL DE BOOR AND KLAUS HÖLLIG 793

Hex Games and Twist Maps on the Annulus . . STEVE ALPERN AND ANATOLE BECK 803

How to Pick Out the Integers in the Rationals:
An Application of Number Theory to Logic DAN FLATH AND STAN WAGON 812

NOTES

A Simple Proof of a Simple Version of the Riemann
Mapping Theorem by Simple Functional Analysis P. R. GARABEDIAN 824

On the Equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$ ALFONSO VILLANI 826

The Ranks of Tournament Matrices D. DE CAEN 829

Convergence-Preserving Functions: An Alternative Discussion ARTHUR SMITH 831

Crossing the Square KEITH WHITTINGTON 833

THE TEACHING OF MATHEMATICS

Lines of Best Fit by Graphics and the Wald Line R. J. DUFFIN 835

On the Mean Value Theorem, Inequality, and Inclusion . . M. FURI AND M. MARTELLI 840

Differentiation of Approximately Specified Functions C. W. GROETSCH 847

More on Cross Products PAUL BINDING 850

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 852

Advanced Problems and Solutions 862

REVIEWS

Pearls in Graph Theory, a Comprehensive Introduction.
By Nora Hartsfield and Gerhard Ringel JOAN P. HUTCHINSON 873

TELEGRAPHIC REVIEWS 876

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $33 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

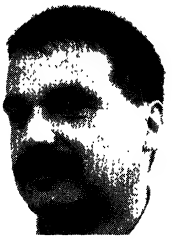
Box-Spline Tilings

CARL DE BOOR[1,2] and KLAUS HÖLLIG[2,3]

CARL DE BOOR has a Ph.D. from the University of Michigan and has been a Professor of Mathematics and Computer Sciences at the University of Wisconsin-Madison since 1972. Piecewise polynomial functions of one or more variables have been his long-term interest.



KLAUS HÖLLIG received his Ph.D. from the University of Bonn in 1979 and is now Professor of Mathematics at the University of Stuttgart. His current research interests are in computer-aided geometric design and related areas of numerical analysis and graphics.



Abstract. We describe a simple method for generating tilings of $\mathbb{R}^d$. The basic tile is defined as

$$\Omega := \{x \in \mathbb{R}^d : |f(x)| < |f(x+j)| \quad \forall j \in \mathbb{Z}^d \setminus 0\},$$

with f a real analytic function for which $|f(x+j)| \to \infty$ as $|j| \to \infty$ for almost every x. We show that the translates of $\overline{\Omega}$ over the lattice $\mathbb{Z}^d$ form an essentially disjoint partition of $\mathbb{R}^d$. As an illustration of this general result, we consider in detail the special case $d = 2$ and

$$f(x) := (\xi^T x)(\eta^T x)$$

with ξ, η column vectors in $\mathbb{Z}^2$. Already this simple choice, which arises in box-spline theory, yields rather interesting partitions of $\mathbb{R}^2$.

Let $f: \mathbb{R}^d \mapsto \mathbb{R}$ be a real analytic function such that, for almost all x and for $j \in \mathbb{Z}^d$,

$$|f(x+j)| \to \infty \quad \text{as } |j| \to \infty. \tag{1}$$

Then the translates of the set

$$\Omega := \Omega(f) := \{x \in \mathbb{R}^d : |f(x)| < |f(x+j)| \quad \forall j \in \mathbb{Z}^d \setminus 0\} \tag{2}$$

provide a tiling for $\mathbb{R}^d$, in the following sense.

THEOREM. *The sets* $\overline{\Omega} + j$, $j \in \mathbb{Z}^d$, *form an essentially disjoint partition of* $\mathbb{R}^d$, *i.e.*

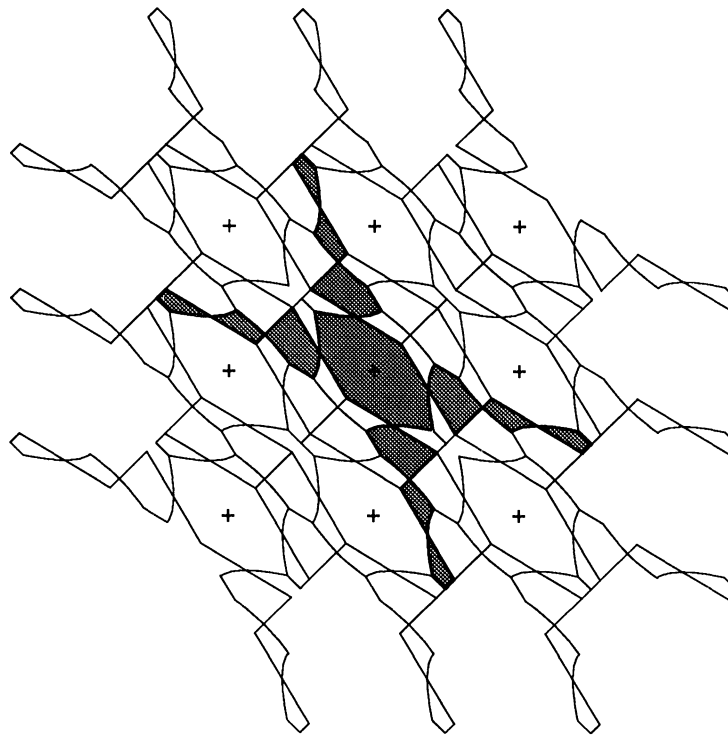
(i) $\overline{\Omega} \cap (\Omega + j) = \varnothing \ \forall j \neq 0$;
(ii) $\text{meas}(\mathbb{R}^d \setminus (\Omega + \mathbb{Z}^d)) = 0$;
(iii) $\text{meas}(\Omega) = 1$.

[1]Supported by the National Science Foundation under Grant No. DMS-8701275
[2]Supported by the United States Army under Contract No. DAAG29-80-C-0041
[3]Sponsored by the National Science Foundation under Grant No. DMS-8351187

Such sets Ω arise in box-spline theory, in the characterization of functions of exponential type as limits of multivariate cardinal series (cf. the Appendix). In that setting, the functions f have the simple form

$$f_{\Xi}(x) = \prod_{\xi \in \Xi} \xi^T x,$$

in which x, ξ are taken to be d-vectors, Ξ is a multiset from $\mathbb{Z}^d \setminus 0$ which spans $\mathbb{R}^d$, and ξ^T denotes the transpose of ξ. Already for $d = 2$ and for Ξ consisting of just two vectors, even these very simple f give rise to surprisingly complex (and strangely beautiful) $\Omega = \Omega_{\Xi}$.



(3) $\overline{\Omega}_{\Xi}$ for $\Xi = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

Proof of the theorem. To prove (i), let $x = \lim x_n$ with $x_n \in \Omega$ and $x - j \in \Omega$. Then, the definition of Ω leads to the contradiction

$$1 > \frac{|f(x-j)|}{|f((x-j)+j)|} = \frac{|f(x-j)|}{|f(x)|} = \lim \frac{|f(x_n - j)|}{|f(x_n)|} \geqslant 1.$$

For the proof of (ii), we deduce from (1) that the function

$$j \mapsto f(x+j)$$

has a minimum for almost all x. If this minimum is unique, then there exists j^* so that

$$|f(x+j^*)| < |f(x+j)| \quad \forall j \neq j^*,$$

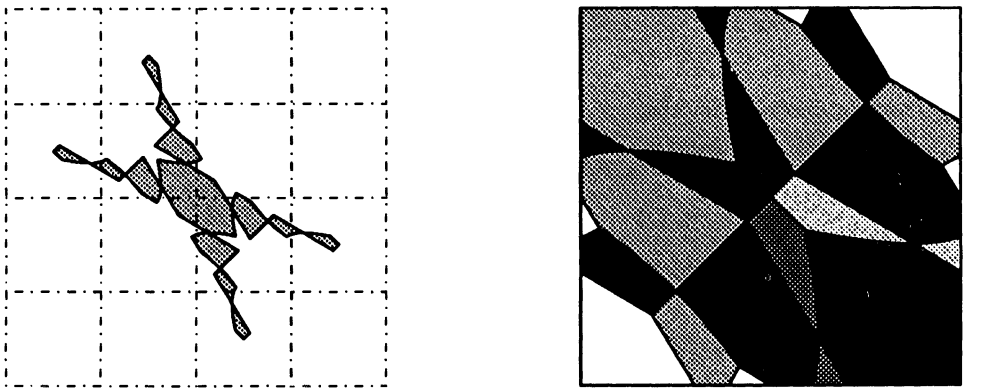
and therefore $x \in \Omega - j^*$. Consequently, up to a set of measure zero, the set $\mathbb{R}^d \setminus (\Omega + \mathbb{Z}^d)$ lies in the union of the zero sets of the (countably many) functions

$$g(x) := |f(x+j)|^2 - |f(x+k)|^2, \qquad j \neq k.$$

Since each such g is analytic, its zero set is of measure zero unless g vanishes identically. But, this latter possibility is excluded since $g = 0$ implies that f is periodic in the direction $j - k$ and this would contradict assumption (1).

For the proof of (iii), we conclude from (i) and (ii) that, up to a set of measure zero, $[0,1]^d$ is the disjoint union of the sets $[0,1]^d \cap (\Omega + j)$ with $j \in \mathbb{Z}^2$, while Ω is the disjoint union of the sets $([0,1]^d - j) \cap \Omega$ with $j \in \mathbb{Z}^2$, and

$$\operatorname{meas}\big([0,1]^d - j\big) \cap \Omega = \operatorname{meas}\big([0,1]^d \cap (\Omega + j)\big).$$



(4) $\Omega_{\left[\begin{smallmatrix}1 & 3\\ 3 & 1\end{smallmatrix}\right]}$ rearranged to fill the unit square.

Special case. In this article, we limit ourselves to the very special case

$$f(x) = (\xi^T x)(\eta^T x), \qquad x \in \mathbb{R}^2,$$

with $\xi, \eta \in \mathbb{Z}^2$ linearly independent.

In this situation, it is convenient to introduce the new variables

$$(u, v) := \Xi^T x = (\xi^T x, \eta^T x).$$

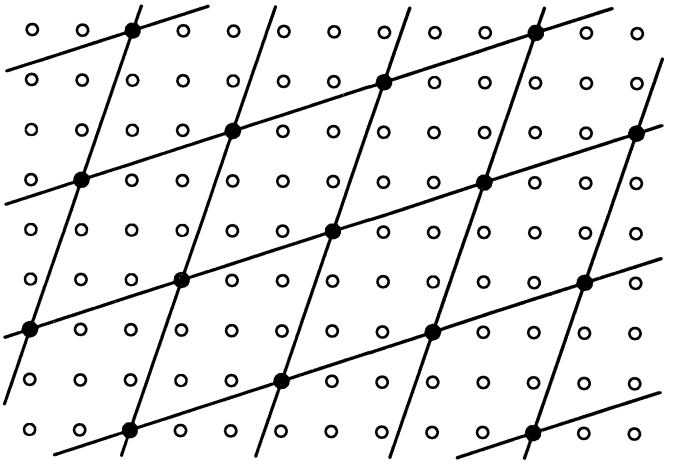
In these new coordinates, the definition of Ω becomes

$$\Omega(\Gamma) := \{(u,v): |u|\,|v| < |u+\alpha|\,|v+\beta| \text{ for } (\alpha,\beta) \in \Gamma \setminus 0\}$$

with

$$\Gamma := \Xi^T \mathbb{Z}^2$$

a sublattice of $\mathbb{Z}^2$.



(5) Sublattice Γ for $\Xi = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

The original Ω can always be recovered via the linear transformation

$$\Omega(\Xi) = (\Xi^T)^{-1}\Omega(\Gamma).$$

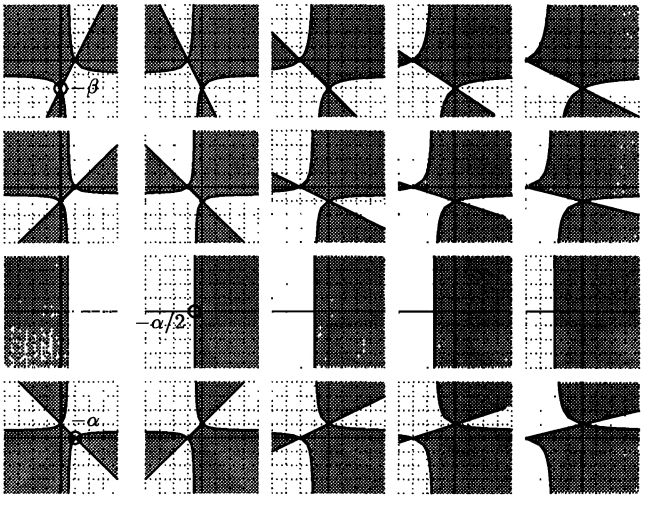
Therefore, in the new coordinates,

$$\text{meas}(\Omega) = |\det \Xi|. \tag{6}$$

Also, the tiling is now obtained by translating Ω over the *sublattice* Γ (rather than over $\mathbb{Z}^2$). On the other hand, we have gained much simplicity since now all possible Ω are intersections of some of the *same* sets $\Omega_{\alpha,\beta}$ with

$$\Omega_{\alpha,\beta} := \{(u,v): |u|\,|v| < |u+\alpha|\,|v+\beta|\}$$

(see (7) FIGURE), different Ω being obtained from different choices of the sublattice Γ.



(7) $\Omega_{\alpha,\beta}$ for $\alpha = -1, 1, 2, 3, 4$ and $\beta = -1, \ldots, 2$.

Symmetries. We now investigate how many essentially different tiles we can obtain in this way. We begin by noting the following obvious symmetries.

(i) Since $\Gamma = -\Gamma$, we also have $\Omega = -\Omega$.

(ii) Γ does not change if Ξ^T is multiplied from the right by a unimodular matrix, i.e. an integer matrix with determinant ± 1.

In particular, we may restrict attention to Ξ^T of the form

$$\begin{bmatrix} p & a \\ 0 & \varepsilon \end{bmatrix} \quad \text{with } p := |\det \Xi|/\varepsilon,\ \varepsilon := \gcd(\eta_1, \eta_2),$$

and $a \in [0, p]$. For, with σ the appropriate sign, $\eta^* := \sigma(\eta_2, -\eta_1)/\varepsilon \in \mathbb{Z}^2$ is carried by Ξ^T to $(\sigma \det \Xi/\varepsilon, 0) = (p, 0)$, while the fact that η_1/ε and η_2/ε are relatively prime implies the existence of an integer vector y for which $\eta^T y = \varepsilon$. Thus, for some choice of the integer c, Ξ^T carries $\gamma := c\eta^* + y \in \mathbb{Z}^2$ to (a, ε) with $a \in [0, p]$. Consequently, $\left[\begin{smallmatrix} p & a \\ 0 & \varepsilon \end{smallmatrix}\right] = \Xi^T[\eta^*, \gamma]$, with $[\eta^*, \gamma]$ necessarily unimodular.

(iii) The scaling

$$\Gamma \mapsto \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix}\Gamma$$

changes Ω correspondingly to

$$\begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix}\Omega.$$

We consider such Ω obtainable one from the other by such scaling as essentially the same. This means that we may further restrict attention to Ξ^T of the form $\Xi^T = \left[\begin{smallmatrix} p & a \\ 0 & 1 \end{smallmatrix}\right]$ with $0 < a < p$ and $\gcd(a, p) = 1$. In fact, since

$$\begin{bmatrix} p & p-a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} p & a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix},$$

it is sufficient to consider Ξ^T of the form

$$\begin{bmatrix} p & a \\ 0 & 1 \end{bmatrix}, \quad \text{with } 0 < a < p/2 \quad \text{and} \quad \gcd(a, p) = 1. \tag{8}$$

In particular, there is just one lattice of interest for each value of $p < 5$, and $p = 7$ is the first value for which there are, offhand, three lattices of interest.

The resulting lattices

$$\Gamma = \Gamma_{p,a} := \begin{bmatrix} p & a \\ 0 & 1 \end{bmatrix}\mathbb{Z}^2, \quad 0 < a < p/2 \cdots < p/2 \quad \text{and} \quad \gcd(a, p) = 1,$$

are indeed different one from the other in that, e.g., $(a, 1)$ is the only point in $\Gamma_{p,a}$ of the form $(b, 1)$ with $0 \leqslant b < p$. This follows from the fact that

$$\min\{b > 0 \colon (b, 0) \in \Gamma_{p,a}\} = p. \tag{9}$$

The corresponding statement

$$\min\{b > 0 \colon (0, b) \in \Gamma_{p,a}\} = p \tag{10}$$

also holds since

$$(\Xi^T)^{-1} = \begin{bmatrix} 1/p & -a/p \\ 0 & 1 \end{bmatrix},$$

hence $(\Xi^T)^{-1}(0, b) = (-ba/p, b)$, and, since $\gcd(a, p) = 1$, this is in $\mathbb{Z}^2$ iff $p|b$.

Bounds. We conclude from (9) and (10) that

$$\Omega \subset \Omega_{\alpha,0} \cap \Omega_{-\alpha,0} \cap \Omega_{0,\alpha} \cap \Omega_{0,-\alpha}$$

with $\alpha = p$. The sets appearing on the right hand side are halfspaces (cf. (7) FIGURE); e.g.

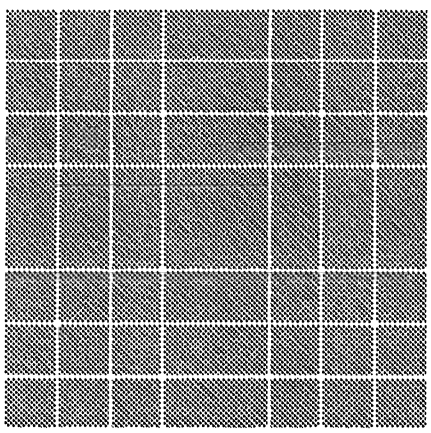
$$\Omega_{\alpha,0} = \{(u,v): u > -\alpha/2\}.$$

Consequently,

$$\Omega \subset (p/2)[-1,1]^2. \tag{11}$$

Note that this bounding square has area p^2, while Ω has area p. This implies that $\Omega = [-1,1]^2/2$ when $p = 1$. It indicates that, for large p, Ω is a rather small subset of this bounding square.

Certain lines are excluded from Ω. Since $|u + \alpha| = 0$ for $u = -\alpha$, Ω cannot contain any point (u,v) with $u = -\alpha$, for which $(\alpha,\beta) \in \Gamma$ for some β. This condition holds for every $\alpha \in \mathbb{Z} \setminus 0$, hence Ω meets none of the lines $u + \alpha = 0$ (therefore also none of the lines $v + \alpha = 0$) for $\alpha \in \mathbb{Z} \setminus 0$.



(12) Ω must lie inside such a set.

We conclude from (11) that, in constructing $\Omega = \bigcap_{j\in\Gamma}\Omega_j$, we only need to consider

$$j \in p[-1,1]^2. \tag{13}$$

For, if $(u,v) \in (p/2)[-1,1]^2$ and, e.g., $(\alpha,\beta) > 0$, then

$$|u + \alpha|\,|v + \beta| < |u + \alpha + mp|\,|v + \beta + np|$$

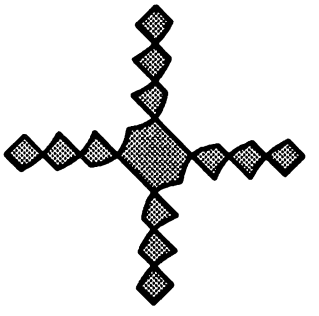
for any positive integers m and n. Consequently

$$x \in (p/2)[-1,1]^2 \cap \bigcap_{j\in\Gamma\cap[0,p]^2} \Omega_j \quad \Rightarrow \quad x \in \bigcap_{j\in\Gamma\cap\mathbb{Z}_+^2} \Omega_j.$$

Figures. We conclude this note with pictures of the first few essentially different tilings obtained in this special case.

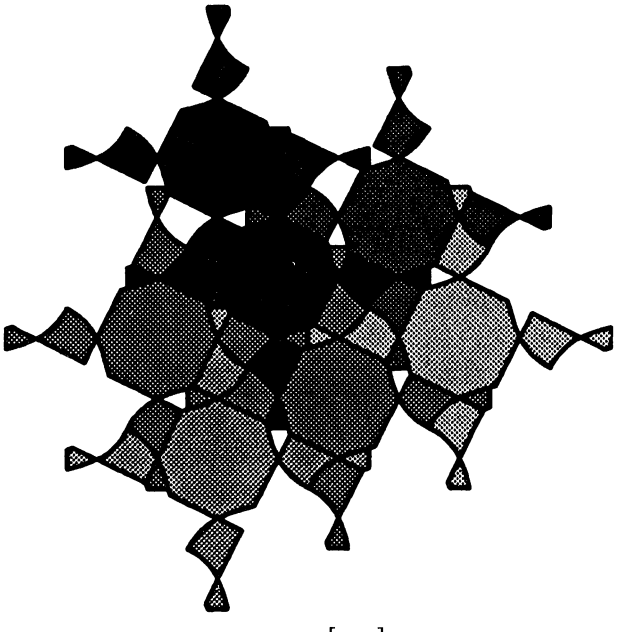
For every p, there is a lattice Γ generated by $(p,0)$ and $(1,1)$, viz. $\Gamma = \Gamma_{p,1}$. For $p = 1$, the corresponding tile is the centered square of side length 1. For $p = 2$, it

is the centered diamond with side length $\sqrt{2}$, i.e., the diamond with vertices at the unit vectors. As p increases, the central portion of the confining set shown in (12) FIGURE is too small to contain all of Ω, and Ω sprouts four arms. The lattice is invariant under the map $(u, v) \mapsto (v, u)$ (in addition to the symmetry $\Gamma = -\Gamma$ observed earlier), hence so is Ω. The resulting four-fold symmetry implies that, in constructing Ω, only one of its four 'arms' need be calculated. The corresponding Ω all look similar, and the following figure gives a typical example.



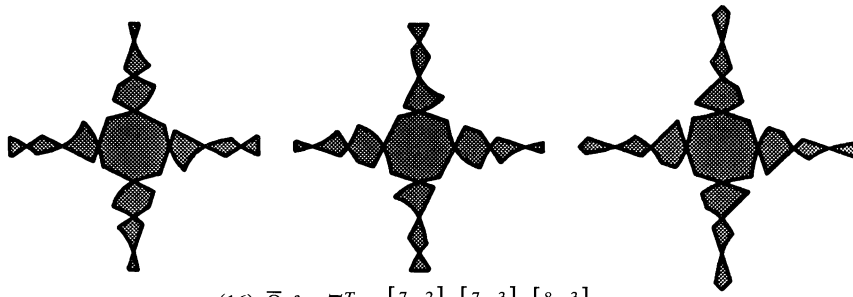
(14) $\overline{\Omega}$ for $\Xi^T = \begin{bmatrix} 8 & 1 \\ 0 & 1 \end{bmatrix}$.

The first tiling of a different kind occurs for $p = 5$. Since its lattice, $\Gamma_{5,2}$, is invariant under rotation of 90°, so is the tile.

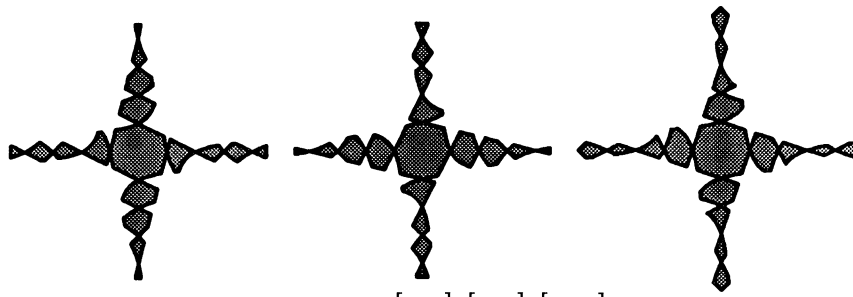


(15) $\overline{\Omega}$ for $\Xi^T = \begin{bmatrix} 5 & 2 \\ 0 & 1 \end{bmatrix}$.

Here are the next few 'unorthodox' tiles.



(16) $\overline{\Omega}$ for $\Xi^T = \begin{bmatrix} 7 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 7 & 3 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 8 & 3 \\ 0 & 1 \end{bmatrix}$.

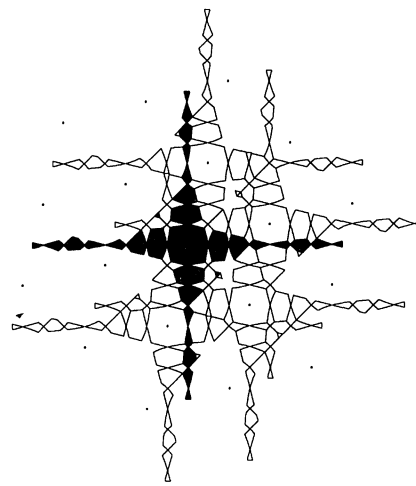


(17) $\overline{\Omega}$ for $\Xi^T = \begin{bmatrix} 9 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 9 & 4 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 10 & 3 \\ 0 & 1 \end{bmatrix}$.

Based on the above figures, one might conjecture that the set Ω is confined to the union

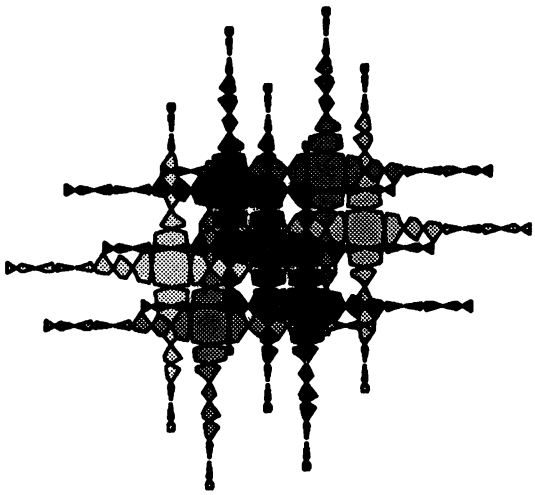
$$[-1,1] \times [-p/2, p/2] \cup [-p/2, p/2] \times [-1,1]$$

of the two central strips of (12) FIGURE. As (18) FIGURE shows, this is in general not true. In fact, rather complicated patterns develop as p increases. The smallest p for which we first encounter a disconnected tile is $p = 15$, and this is the tile shown in (18).



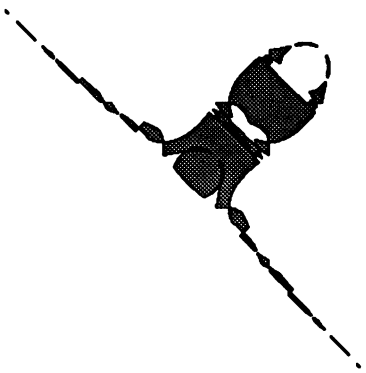
(18) A disconnected tile: $\Xi^T = \begin{bmatrix} 15 & 4 \\ 0 & 1 \end{bmatrix}$.

The next figure shows a more elaborate tile.

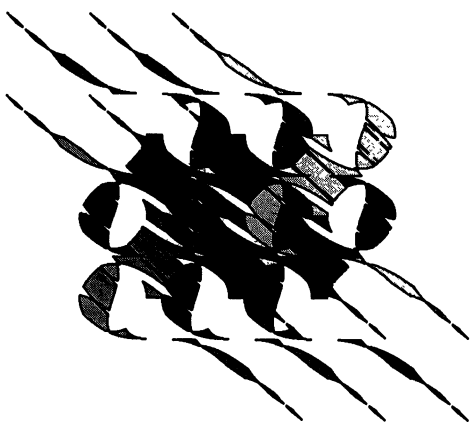


(19) Tiling for $\Xi^T = \begin{bmatrix} 17 & 5 \\ 0 & 1 \end{bmatrix}$.

As we mentioned in the beginning, we have considered in this paper a very special choice of f, motivated by results from box-spline theory. Our final figures give a hint of things to come [$\mathbf{BH}_2$].



(20) The BUG: generating function $f(x, y) := x^3 + y^3 - 2xy$.



(21) NOVA: generating function $f(x, y) := x^3 + y^3 - x$.

Acknowledgement. The authors would like to thank Stefan Pumm and Oliver Schüle from the University of Stuttgart for helping with the figures, and two referees for very helpful comments.

Appendix. We discuss briefly the connection to box-spline theory. Let M: $\mathbb{R}^d \mapsto \mathbb{R}$ be a bounded function with compact support and denote by

$$M_n := M * \cdots * M$$

the n-fold convolution of M. Further, denote by

$$S_n := \left\{ \sum_{j \in \mathbb{Z}^d} M_n(\cdot - j)c(j) : c \in l_2(\mathbb{Z}^d) \right\}$$

the linear span of the translates of M_n with square summable coefficients. We showed in [**BHR**] that a function $g \in L_2(\mathbb{R}^d)$ can be approximated by a sequence $g_n \in S_n$, $n \in \mathbb{N}$, if and only if the support of the Fourier transform of g is contained in the set

$$D(M) := \left\{ x \in \mathbb{R}^d : \left|\widehat{M}(x + 2\pi j)\right| < \left|\widehat{M}(x)\right|, \, j \in \mathbb{Z} \setminus 0 \right\}. \tag{22}$$

With minor modifications, this agrees with the definition of the basic tile in (2), i.e.

$$D(M) = 2\pi\Omega(f), \quad \text{with } f := 1/\hat{M}(2\pi \cdot).$$

Thus the fundamental domain D generates a tiling of $\mathbb{R}^d$.

In the main application of this result, M is chosen as the centered box-spline. Its Fourier transform has the simple form (cf. [$\mathbf{BH}_1$], [**H**])

$$\widehat{M}(x) := \widehat{M\Xi}(x) := \prod_{\xi \in \Xi} \operatorname{sinc}(\xi^T x/2)$$

where $\operatorname{sinc}(t) := \sin t/t$ and Ξ is a multiset of integer d-vectors. Because of periodicity, the factors $\sin(\xi^T x/2)$ are irrelevant for the definition of the fundamental domain, hence

$$D(M_\Xi) = 2\pi\Omega(f_\Xi), \quad \text{with } f_\Xi := \prod_{\xi \in \Xi} \xi^T x.$$

The simplest special case, when $d = 2$ and Ξ consists of just two vectors, is considered in the present paper.

REFERENCES

[BH_1] C. de Boor and K. Höllig, B-splines from parallelepipeds, *J. d'Analyse Math.*, 42 (1982/83) 99–115.

[BH_2] C. de Boor and K. Höllig, Tilings via generating functions, ms.

[BHR] C. de Boor, K. Höllig, and S. D. Riemenschneider, Convergence of Cardinal Series, *Proc. Amer. Math. Soc.*, 98 (1986) 936–939.

[H] K. Höllig, Box-splines, in: *Approximation Theory V*, C. K. Chui, L. L. Schumaker and J. Ward, eds., Academic Press, 1986, pp. 71–95.

Hex Games and Twist Maps on the Annulus

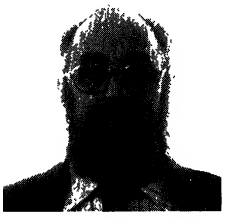
STEVE ALPERN*, *London School of Economics*

ANATOLE BECK, *University of Wisconsin*

STEVE ALPERN studied game theory at Princeton under Oskar Morgenstern and wrote his Ph.D. thesis on ergodic theory with Peter Lax (and unofficially, John Oxtoby) at the Courant Institute. Since 1980 he has been at the London School of Economics.



ANATOLE BECK: I was born in the Bronx and educated at Stuyvesant High School, Brooklyn College, and Yale University, where I syudied with Shizuo Kakutani. After two postdoctoral years at Tulane and in Europe, I joined the faculty of the University of Wisconsin in 1958. I left Wisconsin in 1973 with the intention of spending my remaining years as Professor of Mathematics at the London School of Economics, returning after two years for personal reasons. I have retained my professional and personal ties to the LSE, however, and work there whenever my schedule permits.



1. Introduction. In an elegant paper [G], David Gale has given a combinatorial proof that the game of hex cannot end in a tie, and shown that this fact (and its higher dimensional analog) is equivalent to Brouwer's Fixed Point Theorem. In this paper we define a new version of hex, called either cylindrical or annular hex, by wrapping the standard hex board around a cylinder. It may then, if desired, be flattened out onto an annulus. It turns out that this game, too, cannot end in a tie. This fact is then used to prove a fixed point theorem for twist maps of the annulus. We also give an explicit winning strategy for one of the players. (No such strategy is yet known for standard hex.)

We begin by describing the standard version of the game of hex, which we will call $m \times n$ rectangular hex. Actually only the square game ($m = n$) is ever played, but our analysis requires this greater generality. Two players, called Black and White, alternate in coloring the hexagons of the $m \times n$ rectangular hex board which is pictured (in two possible drawings) in FIGURE 1 with $m = 8$ and $n = 7$. There are two versions corresponding to which player moves first. In any case White wins if he can connect the top to the bottom by white hexagons; Black wins if he can connect the left to the right by black hexagons. Observe that Black has a win in both of the colorings of FIGURE 1.

The $m \times n$ cylindrical hex board is obtained by identifying the right and left boundaries of the rectangular board. This is the same as adding an additional column of hexagons on the right and identifying these hexagons with the ones at the same height on the left column. The cylindrical hex board can also be drawn in the plane, on the annulus (region between concentric circles). The cylindrical

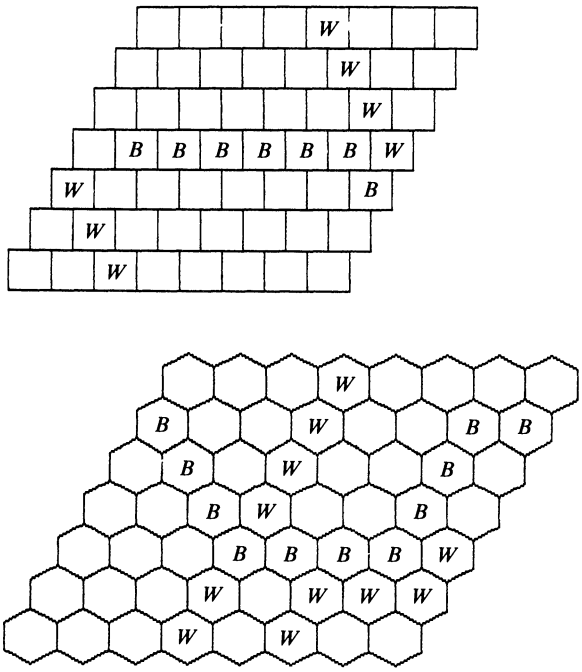


FIG. 1. Two drawings and colorings of the 8 by 7 rectangular hex board.

8×7 hex boards are shown in FIGURE 2, drawn on the annulus. The left and right boundaries of a rectangular hex board have been identified along the thickened line on the annulus going from the inner circle to the outer circle. The colorings of the top and bottom rectangular boards of FIGURE 1 have been wrapped around the cylinder and flattened onto the top and bottom annular boards of FIGURE 2.

In cylindrical hex White wins, as before, if he obtains a path of white hexagons from the top of the cylinder to the bottom. On the annulus this is a white path from the inner circle to the outer circle. For Black to win, he must obtain a circuit of black hexagons which goes around the central axis of the cylinder. On the annulus, this is a circuit with the origin (center of the concentric circles) inside. These definitions implicitly involve the Jordan Curve Theorem but this can be replaced by a purely combinatorial condition. Observe that the coloring of the board on the top in FIGURE 2 is a win for White, but the coloring on the bottom is a win for Black. It is clear that the identification process which creates annular hex from rectangular hex preserves winning paths for White but not necessarily for Black.

In section 2 we prove (Theorem 1) that cylindrical hex cannot end in a tie, and in section 3 we prove (Theorem 2) that when m is even White can always win.

We now consider the question of the existence of fixed points of a homeomorphism f of the annulus. At this point it is easiest to describe the annulus in polar coordinates, as the set of all points in the plane with radial coordinate r between a and b. The obvious example of a fixed point free f is a non-trivial rotation. This is excluded by requiring that f is a "twist." This means that it rotates one boundary circle clockwise and the other, anticlockwise. This condition avoids rotations but still allows fixed point free homeomorphisms of the class $f(\theta, r) = (\theta + c(r), g(r))$

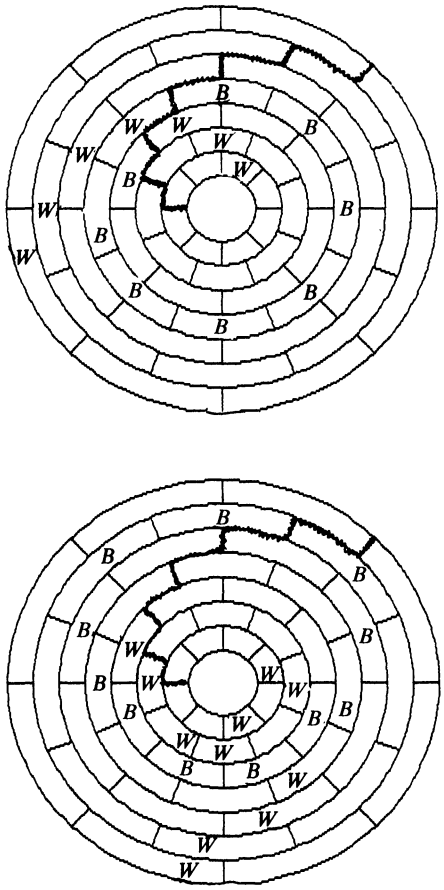


FIG. 2. Two colorings of the 8 by 7 annular hex board.

where c and g are continuous on $[a, b]$, with $c(a) > 0$, $c(b) < 0$, and g strictly increasing on (a, b). Such an example is not area-preserving. Poincare's "Last Geometric Theorem" [**P**], established in full generality by G. D. Birkhoff [**B**], states that an area-preserving twist homeomorphism of the annulus has two fixed points. However we will concentrate on another property of the above family of fixed point free twist maps. They all have what we will call a "rising curve". This is a curve which goes around the annulus, for which every point on it is moved closer to the outer circle. More generally, we will consider "monotone curves" which are either rising curves or falling curves. In section three we use the "No-tie" Theorem (2) for annular hex to prove (Theorem 3) that a twist map of the annulus with no monotone curve must have a fixed point. The conditions of area-preserving and no monotone curve are independent in that in general neither implies the other. Of course for the special case of "monotone twists" the no monotone curve assumption is implied by area preserving. For a twist map with no monotone curve our proof of Theorem 2, which implicitly incorporates Gale's algorithm, can be used to find approximate fixed points.

2. No-tie theorem for cylindrical hex. In order to show that cylindrical hex cannot end in a tie, we give a graph theoretic formulation of both rectangular hex

and cylindrical hex. In this formulation the hexagons pictured in the previous section become the vertices of a graph, and hexagons sharing a common face are adjacent. Thus the hex graphs defined in this section are the dual graphs of the planar graphs of FIGURES 1 and 2.

We begin by defining an "infinite strip" graph S_n. The vertices of S_n are the lattice points $V_n = \{(i, j): i, j \in Z, 1 \leqslant j \leqslant n\}$. Two vertices $z = (z_1, z_2)$ and $w = (w_1, w_2)$ in V_n are adjacent if $|z_1 - w_1| \leqslant 1$, $|z_2 - w_2| \leqslant 1$, and either $z \leqslant w$ (meaning $z_1 \leqslant w_1$ and $z_2 \leqslant w_2$) or $w \leqslant z$. A portion of the infinite planar graph S_n is shown in FIGURE 3. Note that the edges are all either horizontal, vertical, or upward-sloping diagonal.

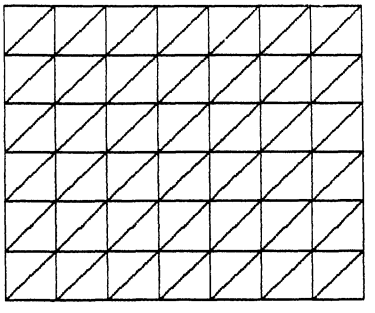


FIG. 3. $H_r(8, 7)$.

The $m \times n$ rectangular hex graph $H_r(m, n)$ is the subgraph of S_n determined by the vertex set $V(m, n) = \{z \in V_n : 1 \leqslant z_1 \leqslant m\}$. To define the $m \times n$ cylindrical hex graph $H_c(m, n)$ with the same vertex set $V(m, n)$ first consider the map ϕ: $V_n \to V(m, n)$ defined by $\phi(z_1, z_2) = (z_1 \bmod m, z_2)$, with $z_1 \bmod m$ chosen in $\{1, \ldots, m\}$. Two vertices z and w in $V(m, n)$ are adjacent in the cylindrical hex graph $H_c(m, n)$ if and only if there are adjacent vertices z' and w' in S_n with $z = \phi(z')$ and $w = \phi(w')$. This means that the cylindrical hex graph $H_c(m, n)$ has all the edges of $H_r(m, n)$ plus the n edges between (m, j) and $(1, j)$ for $j = 1, \ldots, n$ and the $n - 1$ edges between (m, j) and $(1, j + 1)$ for $j = 1, \ldots, n - 1$. FIGURE 3 depicts the 8×7 rectangular hex graph and FIGURE 4 gives a planar (annular) drawing of the 8×5 cylindrical hex graph.

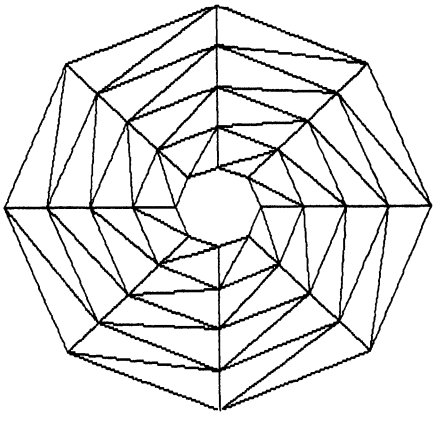


FIG. 4. $H_c(8, 5)$.

Let L: $V(m, n) \to \{\text{black}, \text{white}\}$ be any coloring of the vertices (hexagons) $V(m,n)$. An L-winning path for White in rectangular $m \times n$ hex is any path in $H_r(m, n)$ of white vertices from some z with $z_2 = 1$ to some w with $w_2 = n$. Similarly an L-winning path for Black is any path of black vertices from some z with $z_1 = 1$ to some w with $w_1 = m$. Observe that if a player has an L-winning path, he also has a simple (non self-intersecting) L-winning path.

In cylindrical hex $H_c(m, n)$ an L-winning path for White is similarly any path in $H_c(m, n)$ of white vertices from some z with $z_2 = 1$ to some w with $w_2 = n$. An L-winning path for Black is a bit more difficult to define, but we can give a combinatorial version of a "circuit of black hexagons going around the axis of the cylinder". Let L^*: $V_n \to \{\text{black}, \text{white}\}$ be the "lifting" of the coloring L of $V(m, n)$ to the infinite strip V_n by the formula $L^*(z) = L(\phi(z))$. Observe that the lifted coloring L^* is m-periodic in the horizontal coordinate, since for all (i, j) in V_n we have $L^*(i + m, j) = L(\phi(i + m, j)) = L(\phi(i, j)) = L^*(i, j)$. An L-winning path for Black in the cylindrical hex graph $H_c(m, m)$ can now be defined as the ϕ-image of any path P in S_n of L^*-black vertices between distinct vertices z and w with $\phi(z) = \phi(w)$ (in other words $z_1 \neq w_1$, $z_1 - w_1 = 0 \bmod m$, and $z_2 = w_2$). The path $\phi(P)$ is a circuit in the graph $H_c(m, n)$ which "goes around the axis of the cylinder".

We are now in a position to give a proof of the no-tie theorem for cylindrical hex. This can be proved directly using the Jordan Curve Theorem, but we prefer to give a purely combinatorial proof based on the similar result for rectangular hex, which was established graph theoretically by D. Gale [**G**] (see also [**B, B, C**] for a topological proof).

THEOREM 0. *For any coloring L of the rectangular $k \times h$ hex graph at least one of the players has an L-winning path.*

THEOREM 1. *Let L be any coloring of the cylindrical $m \times n$ hex graph $H_c(m, n)$. Then either White has an L-winning path or Black has an L-winning circuit "going around the axis of the cylinder".*

Proof of Theorem 1 assuming Theorem 0. The coloring L of $V(m, n)$ can be lifted to a coloring L^* of V_n by $L^*(z) = L(\phi(z))$. Let L' be the restriction of L^* to the subset $V(nm + 1, n)$ of V_n. Considering L' as a coloring of the rectangular hex graph $H_r(nm + 1, n)$ we may apply Theorem 0 with $k = nm + 1$ and $h = n$ to assert that at least one player has an L' winning path. We claim that the same player also has an L-winning path or circuit in the cylindrical hex graph $H_c(m, n)$.

To see this first observe that if $P = z^1, \ldots, z^t$ is an L'-winning path for White in $H_r(nm + 1, n)$, then its image $\phi(P)$ is an L-winning path for White in $H_c(m, n)$, since ϕ preserves adjacency. Next suppose that $P = z^1, \ldots, z^t$ is an L'-winning path for Black in $H_r(nm + 1, n)$. Observe that for each of the $n + 1$ numbers $i = 0, \ldots, n$ there is at least one L'-black vertex w^i belonging to the path P and of the form $(im + 1, y_i)$, for some y_i in the n-set $\{1, \ldots, n\}$. Since there are more y_i's than numbers in $\{1, \ldots, n\}$, it follows from the "pigeon-hole principle" that $y_a = y_b$ for some distinct a and b. Hence the ϕ-image of the portion of the path P between w^a and w^b is an L-winning circuit for Black in $HC(m, n)$, since $am + 1 = [\phi(w^a)]_1 \neq [\phi(wb)]_1 = bm + 1, [\phi(w^a)]_1 - [\phi(w^b)]_1 = 0 \bmod m$, and $[\phi(w^a)]_2 = y_a = y_b = [\phi(w^b)]_2$. This establishes the claim and the theorem.

3. A winning strategy for white in annular hex. Since all the variations of hex discussed in this paper are finite games of perfect information which cannot end in a tie, it follows from the standard recursive game tree argument that one of the players must always have a winning strategy. The strategic analysis of these games considers first which player has a win and second what is an explicit winning strategy. For rectangular hex it is known that the player with the closer sides to connect can win, regardless of who goes first, when the dimensions of the board are not equal. That is, White can win if $m > n$, and Black if $m < n$. In this case explicit strategies are known. For the most important case of square hex, $m = n$, the player with the first move has a win. This fact is based on an indirect argument, generally attributed to Nash. However no explicit general strategy is known for White to win at square hex. For a more extensive discussion of the strategic side of rectangular (actually square) hex, see [**B**, **B**, **C**].

In contrast to the situation for rectangular hex, we are able to give an explicit general winning strategy for White in annular hex. It is independent of the relative dimensions or the "advantage" of the first move, but has the curious hypothesis that the circular dimension is even. (We believe the result is true for odd m, but cannot prove it.)

THEOREM 2. *If $m = 2k$ is even, then regardless of n or who has the first move, there is a winning strategy for White in $m \times n$ annular hex. It can be stated explicitly in terms of the* 180 *degree rotation map $T(i, j) = (i + k \bmod m, j)$ of $V(m, n)$ as follows*:

Case 1 (Black moves first). If Black's last move was to color vertex (hexagon) z, then White colors vertex $T(z)$.

Case 2 (White moves first). White plays as in Case 1 whenever possible. If this is impossible, either because it is the first move and z is undefined, or because $T(z)$ is already colored, then color any uncolored vertex.

Proof. We first consider Case 1, where White goes second. In this case the strategy suggested in the theorem is always feasible, since $T(z)$ must be uncolored: If $T(z)$ were already white, then z must have been black prior to Black's move, contrary to assumption. If $T(z)$ were black, then similarly z would have been colored white prior to Black's last move, also a contradiction.

We now show that Black can never attain a winning circuit if White adopts the Case 1 strategy. Hence by the "no-tie theorem for cylindrical hex" (Theorem 1), White will always obtain a winning path.

Suppose that Black obtains a winning circuit γ on the hex graph $H_c(m, n)$ with a winning move z^*. Then, as shown above, the vertex $T(z^*)$ is uncolored at the time the game ends. Color this vertex $T(z^*)$ white (this is no longer to be interpreted as a move in the game, merely an argument in this proof). Since the map T preserves adjacency, $\gamma^* = T(\gamma)$ is a circuit of white vertices which would be a winning circuit for Black if it were colored black rather than white. We now use the fact that the annular hex graph $H_c(m, n)$ is planar and that in particular it can be drawn on the annulus A centered on the origin with inner radius 1 and outer radius 2 by sending the vertex (i, j) to the polar coordinates $\theta = i2\pi/m$ and $r = (n - 2 + j)/(n - 1)$. Hence we may view the circuits γ and γ^* not only combinatorially, but as Jordan circuits in the plane which contain the origin in

their insides. They are disjoint, since otherwise they would have to meet at a vertex, which would then have to be both black and white. The Jordan Curve Theorem thus requires that one of the curves is inside the other. But since the map T preserves the second vertex coordinate j which determines the radius, the two curves γ and $\gamma^* = T\gamma$ have the same maximum radius. Thus the requirement that one is inside the other is impossible (consider the point with maximum radius on the inside curve). It follows that the assumption at the head of this paragraph, that Black obtains a winning circuit when playing against the case 1 strategy of White, is false.

The proof for case 2 is the same, as the additional elements of the strategy make it feasible.

4. Fixed points of twist maps of the annulus. In this section we will show that a twist map of the annulus which has no fixed point must have a monotone (rising or falling) circuit. Before giving the more formal material needed to rigourously define the terms and prove the theorem (3), we now outline the main ideas of the proof.

Consider that a very fine annular hex graph is drawn on the annulus. For any twist map f of the annulus, which has no fixed point, each vertex of the graph will be moved to another point (not necessarily a vertex) of the annulus. We label the vertex according to which direction it is moved. If it moves mainly out towards the outer boundary label it O; if it moves mainly in towards the inner boundary label it I. If it moves mainly clockwise, label it C; if anti-clockwise, label it A. Thus every vertex can be labeled with either O, I, C, or A. Since f is a twist map, all vertices on the outer boundary will be labeled A, and all vertices on the inner boundary will be labeled C. Now color all vertices labeled A or C white, and color all vertices labeled O or I black. Apply Theorem 1 to assert that there is either a path of white vertices going from the inner to the outer boundary of the annulus, or a circuit of black vertices going around the annulus. Since the map f is continuous, it follows that for a grid size (distance between adjacent vertices) sufficiently small, adjacent vertices cannot be moved in opposite directions. That is, adjacent vertices cannot be labeled A and C, or I and O.

Now suppose that there is a winning path of white vertices, going from an inner boundary vertex labeled C, to an outer boundary vertex labeled A. Somewhere along this path a vertex labeled C would have to be adjacent to a vertex labeled A —but this has been shown to be impossible. So it follows that there is a circuit of black (that is, I or O) vertices going around the annulus. By the same argument about adjacent vertices not having opposite labels, it follows that this circuit is either labeled all I (inwards) or all O (outwards). Suppose it is labeled all I, so that every vertex in the path moves outwards. If the grid size is sufficiently small with respect to the minimum positive distance a point must move under f, then any point on the edge between two of the outward-moving vertices of the black circuit must also be outward-moving. Hence if the combinatorial circuit of black vertices is viewed as a simple closed curve, it is the rising curve whose existence is asserted in Theorem 3. Our proof of Theorem 3 is merely an elaboration of the argument sketched here.

In order to formalize the above arguments we will first of all use the cylindrical form of the annulus, that is, $A = S^1 \times [0, 1] = \{(\theta, r)\colon \theta \text{ in } R/Z,\ 0 \leqslant r \leqslant 1\}$. Furthermore we will "lift" all arguments to the universal covering space S of A,

the horizontal strip $S = \{(x, y)\colon x \text{ in } R,\ 0 \leqslant y \leqslant 1\}$, via the covering map $\pi\colon S \to A$ given by $\pi(x, y) = (x \bmod 1, y)$. A homeomorphism $F\colon S \to S$ is called a lift of $f\colon A \to A$ if $f\pi = \pi F$. All lifted concepts will be invariant with respect to the translation $T(x, y) = (x + 1, y)$ of S. In particular, any homeomorphism $f\colon A \to A$ can be lifted to a homeomorphism $F\colon S \to S$ with $FT = TF$ (which is defined up to a power of T) and any such F is the lift of an annulus homeomorphism f. For z in S write $F(z) = (F_1(z), F_2(z))$, and $\|z\| = \max(|z_1|, |z_2|)$.

DEFINITION. *A homeomorphism $f\colon A \to A$ is called a twist homeomorphism if it preserves orientation and some lift $F\colon S \to S$ satisfies $F_1(x, 0) > x$ and $F_1(x, 1) < x$ for all x in R. In this case we will also call F a periodic twist homeomorphism of S.*

PROPOSITION 1. *Let $F\colon S \to S$ be a periodic twist homeomorphism of the horizontal strip S. Then for any positive number M there is a simple polygonal curve $\gamma\colon R \to S$, with $\gamma(t + 1) = T(\gamma(t))$ for t in R, such that $|F_1(z) - z_1| < M$ for every point $z = \gamma(t)$ on the curve γ.*

Proof. Since F is continuous and commutes with T, it is uniformly continuous. Hence, there is a δ, $0 < \delta < M/4$, such that $\|z - w\| < \delta$ implies that $\|F(z) - F(w)\| < M/4$. Choose n so that $n > 1 + 1/\delta$. Embed the vertex set $V_n = \{(i, j)\colon 1 \leqslant j \leqslant n\}$ into the horizontal strip S via the map $\alpha(i, j) = (i/n, (j - 1)/(n - 1))$. This is the "lifted" analog of drawing the hex graph on the annulus. Observe that if two vertices v and v' are adjacent in V_n then $\|\alpha(v) - \alpha(v')\| \leqslant 1/(n - 1) < \delta$. This is our formalization of a "sufficiently fine hex graph". Define $p\colon S \to R$ by $p(z) = F_1(z) - z_1$ and observe that p is periodic in that $pT = p$. Let L^* be the coloring of the vertex set V_n according to the rule that v is black if and only if $|p(\alpha(v))| < M/2$. Observe that $L^*(i + n, j) = L^*(i, j)$ is a periodic coloring of V_n and thus lifts some coloring L of the annular hex graph $H_c(n, n)$.

By Theorem 1, either White has an L-winning path in $H_c(n, n)$, or Black has an L-winning circuit. First suppose White has a winning path. Let $v^1, \ldots, v^k$ be a path of L-white vertices with $(v^1)_2 = 1$ and $(v^k)_2 = n$. Since F is a twist map it follows that $p(\alpha(v^1)) > 0$ and $p(\alpha(v^k)) < 0$, and hence for some consecutive white vertices v and v' in the path we must have that $p(\alpha(v)) > 0$ and $p(\alpha(v')) < 0$, $\|\alpha(v) - \alpha(v')\| < \delta$. Setting $z = \alpha(v)$ and $w = \alpha(v')$, we have that $\|z - w\| < \delta$ and hence $\|F(z) - F(w)\| < M/4$. Also $p(z) - p(w) \geqslant (M/4) - (-M/4) = M/2$, since z and w are α-images of white vertices. However this last fact is contradicted upon observing that

$$\begin{aligned} |p(z) - p(w)| &= |F_1(z) - z_1 - F_1(w) + w_1| \\ &\leqslant |F_1(z) - F_1(w)| + |w_1 - z_1| \\ &\leqslant \|F(z) - F(w)\| + \|w - z\| \\ &\leqslant M/4 + \delta < M/2. \end{aligned}$$

Since White cannot have an L-winning path, Black must have an L-winning circuit. By definition (see section 2), this is the ϕ-image of an L^* black path $v^1, \ldots, v^k$ in V_n with $\phi(v^1) = \phi(v^k)$ and $v^1 \neq v^k$. Let $\gamma^*(t)$, $0 \leqslant t \leqslant 1$, be a parameterization of the polygonal path on S which connects the α-image of these vertices, and let $\gamma\colon R \to S$ be the extension of γ^* satisfying $\gamma(t + 1) = T(\gamma(t))$. Now let $z = \gamma(t)$ be any point on the circuit γ. For some i, z is on an edge of the polygonal path γ with a vertex $w = \alpha(v^i)$ at its end. Hence $\|z - w\| < \delta$ for a

black vertex w satisfying $|p(w)| < M/2$. The proposition now follows from the estimate

$$\begin{aligned} |p(z)| &= |F_1(z) - z_1| \\ &= |F_1(z) - F_1(w) + F_1(w) - w_1 + w_1 - z_1| \\ &\leqslant \|F(z) - F(w)\| + |p(w)| + \|w - z\| \\ &< M/4 + M/2 + M/4 = M. \end{aligned}$$

THEOREM 3. *Let $f\colon A \to A$ be any twist map of the annulus (cylinder) $A = \{(\theta, r)$: θ in R/Z, $0 \leqslant r \leqslant 1\}$ which has no fixed point. Denote $(\theta', r') = f(\theta, r)$. Then there is a simple closed curve P which goes around the axis of the cylinder, with either*

(i) *$r' > r$ for all points (θ, r) on P (P is a rising curve), or*

(ii) *$r' < r$ for all points (θ, r) on P (P is a falling curve).*

Furthermore, given any $\beta > 0$ we may choose P so that it rises or falls arbitrarily steeply, in that $|\theta' - \theta|/|r' - r| < \beta$. Hence any twist map of the annulus without arbitrarily steep monotone curves has a fixed point.

Proof. Let $F\colon S \to S$ be a twist map which lifts f. If z' were fixed by F then $f\pi(z') = \pi F(z') = \pi(z')$ so f would also have a fixed point, contrary to assumption. Hence F has no fixed points. Since F is continuous and $FT = TF$ it follows that F is uniformly continuous. The last two sentences together imply the existence of a positive number D such that $\|F(w) - w\| > D$ for all w in S. Choose a number M such that $M < \beta D$ and $M < D$. By Proposition 1 there is a curve γ on S with $\gamma(t + 1) = T(\gamma(t))$ such that $|p(z)| < M$ for all z on γ, where $p(w) = F_1(w) - w_1$. Define $q(w) = F_2(w) - w_2$. Since $D < \|F(w) - w\| = \max[|p(w)|, |q(w)|]$ for all w in S, it follows that $|q(z)| > D$ for all z on γ. Since γ is a connected set it follows that either q is positive on all of γ or q is negative on all of γ. Without loss of generality we may assume q is positive. In this case $P = \pi\gamma$ is a rising curve for f on A, since for any z on γ we have $F_2(z) > z_2$ and hence for any point $(\theta, r) = \pi(z)$ on P we have that $r' = F_2(z) > z_2 = r$. Furthermore $|\theta' - \theta| = |p(z)| < M$ so that $|\theta' - \theta|/|r' - r| < M/D < \beta$, by choice of M.

REFERENCES

[B,B,C] Anatole Beck, Michael N. Bleicher, and Donald W. Crow, Excursion into Mathematics, Worth, New York, 1969, pp. 327–339.

[B] G. D. Birkhoff, An extension of Poincaré's Last Geometric Theorem, *Acta Math.*, 47 (1924) 297–311.

[G] David Gale, The game of hex and the Brouwer fixed point theorem, *American Math. Monthly*, 86 (1979) 818–827.

[P] H. Poincaré, Le problem de trois corps et les equations de la dynamique, *Acta Math.*, 13 (1890) 1–271.

How to Pick Out the Integers in the Rationals: An Application of Number Theory to Logic

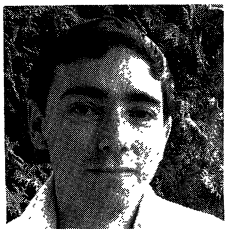
DAN FLATH, *University of South Alabama, Mobile, AL 36688*

STAN WAGON, *Macalester College, St. Paul, MN 55105*

DAN FLATH is Assistant Professor of Mathematics at the University of South Alabama. His principal mathematical interests are number theory and Lie groups. He earned his Ph.D. (1977) at Harvard University under Prof. John Tate and has since held positions at Rutgers, the National University of Singapore, Duke University, and the University of Paris.



STAN WAGON received his Ph.D. from Dartmouth College under the direction of James E. Baumgartner and taught at Smith College from 1975 to 1990, with visits to Carleton College, the University of Colorado, the University of California at Berkeley, and the University of Washington. His introduction to the area of Diophantine definitions came in a memorable course taught by Julia Robinson at the University of California, Berkeley, in 1975.



1. Introduction. Everyone is by now familiar with Gödel's famous incompleteness theorem for the natural numbers, $\mathbb{N}$. If any computable list of true statements about $\mathbb{N}$ is taken as a set of axioms (for example, Peano's Axioms), then the set of all statements that can be derived from these axioms by the usual rules of proof is incomplete; that is, there is a true statement about $\mathbb{N}$ that is not provable. A closely related result is that there is no algorithm that decides whether an arbitrary sentence (in the appropriate language) is true in $\mathbb{N}$.

Is the situation any better for the rationals, $\mathbb{Q}$? One way to approach this question is to explore the logical relationship between $\mathbb{N}$ and $\mathbb{Q}$. If it is tight enough then perhaps incompleteness and undecidability for $\mathbb{Q}$ can be deduced from the corresponding results for $\mathbb{N}$. That this program does, in fact, work out was discovered by Julia Robinson in her 1948 dissertation [**13**] written under the direction of Alfred Tarski. Her work—a subtle combination of number theory and logic—yields a result that confirms a strong logical connection between the integers and the rationals. This paper is devoted to an exposition of the details of her discovery.

The essence of the problem is the definability of sets in terms of other sets and operations. An instructive example is whether positivity is definable in the arithmetic of the integers, $\mathbb{Z}$. Precisely: Given the operations of addition and multiplication on $\mathbb{Z}$, and the usual notions of first-order logic ($=$, $\forall$, $\exists$, $\vee$, $\wedge$, $\neg$, $\Rightarrow$, $\Leftrightarrow$), is there a formula $S(X)$ that is true in $\mathbb{Z}$ precisely when X is a nonnegative integer? The reader might like to ponder this before reading on. To appreciate the difficulty, note that it is a simple matter to define any individual integer. If $S_0(X)$

is the formula $X = X + X$, and $S_1(X)$ is $(X = X \cdot X) \wedge \neg(X = X + X)$, then, in $\mathbb{Z}$, S_0 is true only for $X = 0$ and S_1 is true only for $X = 1$. And if $S_n(X)$ is then defined to be the sentence $\exists Y(S_1(Y) \wedge X = Y + Y + \cdots + Y)$ (n summands), it is clear that $S_n(X)$ defines n. Gathering these infinitely many formulas into the single formula $S_0(X) \vee S_1(X) \vee \cdots \vee S_n(X) \vee \cdots$ will work to define the natural numbers in $\mathbb{Z}$, but infinite conjunctions are not part of first-order logic.

Yet the sought-after definition exists. Let $S(X)$ be the formula: $\exists A \exists B \exists C \exists D(X = A \cdot A + B \cdot B + C \cdot C + D \cdot D)$. Because of Lagrange's theorem that every element of $\mathbb{N}$ is a sum of four squares, it is immediate that for $t \in \mathbb{Z}$, $S(t)$ is true in $\mathbb{Z}$ if and only if $t \in \mathbb{N}$. In short, $\mathbb{N}$ is a *definable* subset of $\mathbb{Z}$. This shows the close connection between nontrivial number theory and questions of definability. Using deeper results of number theory one can give simpler definitions, that is, definitions using fewer variables (see §5). On the other hand, the four-squares definition is as simple as possible in the sense that only existential quantifiers are used. By the way, it is clear that not all subsets of $\mathbb{Z}$ are definable: there are uncountably many of them and only countably many possibilities for definitions. A more important reason is due to Tarski who proved that $\{n \in \mathbb{N}$: the sentence with Gödel number[1] n is true in $\mathbb{N}\}$ is not a definable subset of $\mathbb{N}$ (or of $\mathbb{Z}$). In short, truth is not definable. Every recursively enumerable set is definable (essentially due to Gödel), a result that was dramatically improved in 1970, when Matijasevič showed that every r.e. set is, in fact, Diophantine (i.e., definable using only existential quantifiers) (see [**2**], [**9**], and also §5 of this paper; the definable sets as defined here are usually called the *arithmetical* sets, for more on which see [**4**, §3.5] and [**5**]).

Now, what happens when we move to the rationals (again, with addition and multiplication as the primitive operations)? It is an easy exercise (use the sum of four squares) to show that the positive rationals are definable in $\mathbb{Q}$. But are the integers definable in $\mathbb{Q}$? This is one of the problems studied by Robinson in her thesis, and her solution is extremely elegant (and important—an application is given at the end of §2; see also §5). Note that, again using a four-squares argument, $\mathbb{N}$ is definable in $\mathbb{Q}$ if and only if $\mathbb{Z}$ is definable in $\mathbb{Q}$; we focus on the definability of $\mathbb{Z}$. Robinson's result (if not its proof) is well known to logicians. Our purpose here is to give an exposition of her proof with more details than are given in [**13**], in the belief that a wider audience will enjoy the interplay between number theory and elementary logic.

2. The definition of $\mathbb{Z}$. It is not at all obvious how to proceed, assuming one has faith that a definition exists. There is no clear way to get at the denominators of a rational number, which can only be referred to by a variable-name. It might be impossible, or at least very difficult, to use first-order logic to uncover the hidden numerator and denominator of an arbitrary rational. Of course, if one *could* get at the exact denominator (i.e., the lowest-terms denominator), then checking whether it equals 1 would solve the problem. Some hope was provided by an argument Julia Robinson found that yields a definition of those rationals whose exact denominator is not divisible by 2. But the argument is nontrivial and there is no clear way to extend it to the other primes. First of all, one cannot even say what a prime is,

[1] Gödel's revolutionary work in logic depended on the assignment of numbers to formulas of logic, thus enabling him to construct sentences that, when interpreted in $\mathbb{N}$, asserted their own unprovability.

since that would assume knowing what an integer was. And even if a way around this difficulty could be found and formulas guaranteeing p-free denominators did exist, one could not simply gather all the formulas into a single large conjunction. Thus some sort of uniform solution to the problem of clearing the denominators of primes would be needed.

The argument for destroying even denominators is not essential to the main theorem, but is worth examining. It is based on the Gauss–Legendre three-squares theorem for integers, or rather, its extension to the rationals, which we now state.

THEOREM 1. *A nonnegative integer n is a sum of three rational squares if and only if n does not have the form $4^a(8k + 7)$.*

Theorem 1 is much more difficult than the four-squares theorem; the integer version (see, e.g., [**6**] or [**10**] for a proof) states that the only integers that are not a sum of three integer squares are those of the form $4^a(8k + 7)$. The rational version is an easy consequence of the integer version. However, a more modern approach is to first derive Theorem 1 as a straightforward consequence of the Hasse–Minkowski Theorem (the argument is given at the end of §3). Then the integer version follows by a result of L. Aubry (see [**24**, Chap. 3, App. 2]), who, in 1914, gave a beautiful geometric proof that an integer that is a sum of three rational squares is necessarily a sum of three integer squares.

Julia Robinson used Theorem 1 to show that the formula[1]

$$\exists A \exists B \exists C(7X^2 + 2 = A^2 + B^2 + C^2)$$

defines those rationals whose exact denominator is odd. To see that this definition works, first note that for a rational $x = n/d$, $7x^2 + 2$ is a sum of three rational squares iff $7n^2 + 2d^2$ is a sum of three rational squares. Now, if d is odd then $7n^2 + 2d^2$ is congruent to 1 (mod 8) if n is odd, or 2 or 6 (mod 8) if n is even. In either case, Theorem 1 applies to $7n^2 + 2d^2$, which is therefore a sum of three rational squares. And if d is even then $7n^2 + 2d^2 \equiv 7 \pmod 8$ whence $7n^2 + 2d^2$ is not a sum of three squares. Therefore $7x^2 + 2$ is a sum of three rational squares if and only if d is odd.

To further motivate the positive solution to the problem of defining $\mathbb{Z}$ in $\mathbb{Q}$, we point out another illegal way to define the integers. Consider the formula $\Gamma(X)$:

$$\forall A \subseteq \mathbb{Q}\{[0 \in A \wedge \forall R(R \in A \Rightarrow R + 1 \in A)] \Rightarrow X \in A\}.$$

An easy induction shows that for each $n \in \mathbb{N}$, $\mathbb{Q} \vDash \Gamma(n)$ (meaning: $\Gamma(n)$ is true in $\mathbb{Q}$). Conversely, if $\mathbb{Q} \vDash \Gamma(t)$ then, because $A = \mathbb{N}$ satisfies the hypothesis of Γ (i.e., induction is true in $\mathbb{N}$), $t \in \mathbb{N}$. Thus $\Gamma(X)$ defines $\mathbb{N}$ in $\mathbb{Q}$. Of course, this definition does not belong to first-order logic, but to second-order logic; in first-order logic one cannot quantify over subsets.

We can now state Julia Robinson's definition of $\mathbb{Z}$ in $\mathbb{Q}$. Let $S(Q, P, N)$ be

$$\exists A \exists B \exists C(2 + QPN^2 = A^2 + QB^2 - PC^2).$$

Think of Q and P as being fixed, so that $S(Q, P, N)$ defines a subset of $\mathbb{Q}$ via the variable N. Now let $R(X)$, the formula that defines $\mathbb{Z}$, be:

$$\forall Q \forall P\big([S(Q, P, 0) \wedge \forall T\{S(Q, P, T) \Rightarrow S(Q, P, T + 1)\}] \Rightarrow S(Q, P, X)\big).$$

[1]The use of "2" in a formula is allowed because $2 = 1 + 1$ and 1 is easily defined. Precisely, we may recast the formula as: $\exists U \exists A \exists B \exists C((U^2 = U) \wedge (U + U \neq U) \wedge (7X^2 + U + U = A^2 + B^2 + C^2))$.

Exactly as for the second-order definition using $\Gamma(X)$, induction yields that for any $n \in \mathbb{N}$, $\mathbb{Q} \models R(n)$. Moreover, $S(Q, P, N)$ is equivalent to $S(Q, P, -N)$, which implies that for any $n \in \mathbb{Z}$, $\mathbb{Q} \models R(n)$.

It remains to show where the formula $S(Q, P, N)$ comes from and why the truth of $R(x)$ in $\mathbb{Q}$ means that x is an integer. As alluded to above, the idea is that the truth of $R(x)$ forces x's exact denominator to be free of all primes.

The definition of $S(Q, P, N)$ is motivated by two lemmas that Julia Robinson discovered, lemmas whose proofs are not elementary.

LEMMA 1. *If x is rational and p is a prime with $p \equiv 3 \pmod 4$, then $A^2 + B^2 - pC^2 = px^2 + 2$ is solvable in rational A, B, C (i.e., $\mathbb{Q} \models S(1, p, x)$) if and only if the exact denominator of x is odd and not divisible by p.*

LEMMA 2. *If x is rational and p and q are odd primes with $p \equiv 1 \pmod 4$ and $(q/p) = -1$, then $A^2 + qB^2 - pC^2 = qpx^2 + 2$ is solvable in rational A, B, C (i.e., $\mathbb{Q} \models S(q, p, x)$) if and only if the exact denominator of x is relatively prime to p and q.*

Now, assuming these two lemmas (which are proved in §3), we can complete the proof that $R(X)$ defines $\mathbb{Z}$ in $\mathbb{Q}$. Suppose $R(x)$ is true for some rational x and p is a prime congruent to 3 (mod 4). Then $S(1, p, 0)$ holds (by Lemma 1, since the exact denominator of 0 is 1; alternatively, just let $A = B = 1$, $C = 0$). Moreover, for any $m \in \mathbb{Q}$, if $S(1, p, m)$ is true then so is $S(1, p, m + 1)$ (use Lemma 1 and the fact that m and $m + 1$ have the same exact denominators). Therefore the truth of $R(x)$ implies that $S(1, p, x)$ is true, which, again by Lemma 1, implies that x's exact denominator is not divisible by p or by 2. It remains to show that this denominator is not divisible by p when $p \equiv 1 \pmod 4$. First choose q to be an odd prime that is a quadratic nonresidue mod p; such must exist, for if s is any quadratic nonresidue, then one of s, $p + s$ is odd, and hence has an odd prime factor that is a quadratic nonresidue. Then use Lemma 2 to show that $S(q, p, 0)$ holds and that $S(q, p, m)$ implies $S(q, p, m + 1)$ for any rational m. A final application of Lemma 2 to x yields that x's exact denominator is not divisible by p. Therefore the exact denominator of x is 1, completing the proof that $R(X)$ defines $\mathbb{Z}$.

How did Julia Robinson discover this definition? Here is Raphael Robinson's recollection: "I knew the three-square theorem and helped Julia find how to eliminate the factor 2 from the denominator of a rational number. But the three-square theorem exhausted my knowledge of ternary quadratic forms, and Julia knew nothing about them. She started searching the literature for relevant material. She had no idea that the theory was simpler over $\mathbb{Q}$ than over $\mathbb{Z}$. She looked at a lot of things that were not helpful. It was several months before she found the Hasse paper. Then she had to find suitable forms to eliminate the various primes from the denominators. Note that it is only the fact that ternary forms represent most numbers with a few exceptions that makes the definition possible. Binary forms represent too few numbers, quaternary forms too many. I remember asking her what good it would do to find formulas for eliminating all of the primes, if these could not be combined into one formula anyway. But she also found how to do this. The proof would have been a lot easier for someone who already knew about Hasse's work. But I guess that those who knew it had never heard of Tarski's problem. It must often happen that the tools for solving a problem are known, but not to the people working on the problem."

We now discuss briefly the consequence of the definability of $\mathbb{Z}$ in $\mathbb{Q}$ that motivated the study of the problem. A *theory* is a set of sentences in the language of first-order logic with some relation and function symbols. For example, the theory of the natural numbers is usually taken to mean the set of those first-order sentences involving $+$ and $\cdot$ that are true in the structure $(\mathbb{N}, +, \cdot)$. Church (see [**4**, p. 231]) proved that the theory of the natural numbers is *undecidable*, that is, there is no algorithm that decides, on input an arbitrary sentence, whether that sentence is in the theory (i.e., is true in $(\mathbb{N}, +, \cdot)$). Because $\mathbb{N}$ is definable in $\mathbb{Z}$, it is not hard to see that the theory of $(\mathbb{Z}, +, \cdot)$ is undecidable too. Similarly, Julia Robinson's definition of $\mathbb{Z}$ in $\mathbb{Q}$ implies that the theory of $(\mathbb{Q}, +, \cdot)$ is undecidable. This is in contrast to Tarski's work showing that the theories of $\mathbb{R}$, $\mathbb{C}$, $\mathbb{A}$ (the algebraic numbers), and $\mathbb{A} \cap \mathbb{R}$ are decidable (all with the operations $+$ and $\cdot$). A consequence is that $\mathbb{Z}$ is not definable in any of these fields. In fact, Tarski's work yields a complete description of the definable sets in $(\mathbb{R}, +, \cdot)$: A set is definable in this structure iff it is a finite union of intervals with algebraic or infinite endpoints (see [**15**]).

There are several noteworthy extensions of Julia Robinson's theorem. In [**14**] she proved that $\mathbb{Z}$ is definable in any *number field*, that is, any field extension L of the rationals having finite degree. This was proved in two steps. First she gave a definition of $\mathbb{Z}$ in $\mathcal{O}_L$, the ring of algebraic integers in L (the definition is relatively simple); then a definition of $\mathcal{O}_L$ in L is given, a definition that, as in the case of defining $\mathbb{Z}$ in $\mathbb{Q}$, relies on a substantial amount of algebraic number theory. It is still open whether the field of constructible numbers is decidable. J. Denef (see [**3**]) provided a dramatic improvement in the case of quadratic fields ($L : \mathbb{Q} = 2$); he showed that $\mathbb{Z}$ is *Diophantine* over $\mathcal{O}_L$, which means that there is a definition of $\mathbb{Z}$ using only existential quantifiers (the four-squares definition shows that $\mathbb{N}$ is Diophantine over $\mathbb{Z}$; for more on this notion, see §5). Denef and Lipshitz have conjectured that $\mathbb{Z}$ is Diophantine over $\mathcal{O}_L$ for any number field L; this has been proved for several classes of number fields (totally real number fields (Lipshitz), quadratic extensions of totally real number fields (Denef and Lipshitz), Abelian extensions of $\mathbb{Q}$ (A. Shlapentokh and H. N. Shapiro [**20**]). However, L. van den Dries [**23**], building upon work of R. Rumely, showed recently that the ring of all the algebraic integers is decidable, and so the integers are not definable in this ring.

Theories also arise as the set of sentences that can be proved from certain axioms by rules of logic. For example, the elementary theory of groups is the set of those sentences in the language with one operation symbol (the group operation) and one constant symbol (the identity) that are provable from the group axioms. The undecidability of $\mathbb{Z}$ allows one to deduce the undecidability of the theory of groups and of the theory of rings (see [**22**, p. 136]); the main point is that the theories of groups and rings are contained in the theory of $\mathbb{Z}$. It is a consequence of the definability of $\mathbb{Z}$ in $\mathbb{Q}$ that the theory of fields is undecidable. This result is striking because of results of Tarski that the theories of real closed fields and of algebraically closed fields of characteristic zero are complete, which means that these theories coincide with the theories of $\mathbb{R}$ and $\mathbb{C}$, respectively; thus these theories are decidable (see [**15**], [**12**, §§1, 2], [**22**, §5.5]).

One can also look at larger fields. For example, Raphael Robinson has shown [**17**] that the elliptic curve $y^2 = x^3 - 4$ can be used to define $\mathbb{Q}$ in any pure

transcendental extension of $\mathbb{Q}$. This combines with Julia Robinson's work to yield a definition of $\mathbb{Z}$ in such fields.

3. The Hasse–Minkowski Theorem. We say that a quadratic form

$$f(X_1, \ldots, X_r) = \sum_{i=1}^{r} a_i X_i^2$$

with coefficients a_i in a field K *represents zero in* K if and only if the equation $f(X) = 0$ has a solution with $X \in K^r$ and $X \neq \mathbf{0}$.

The nontrivial number theory called up by Julia Robinson can all be described as a series of statements that various rational quadratic forms in four variables do or do not represent zero in $\mathbb{Q}$. These statements are by now routine special cases of one of the great theorems of twentieth-century mathematics, the Hasse–Minkowski Theorem. Our goal in this section is to present an elementary statement of the Hasse–Minkowski Theorem and to show how Robinson's Lemmas 1 and 2 can be deduced from it.

We focus on the case $K = \mathbb{Q}$. Clearing denominators, we henceforth consider only integral ($a_i \in \mathbb{Z}$) quadratic forms $f(X) = \sum_{i=1}^{r} a_i X_i^2$. If f represents 0 in $\mathbb{Q}$ then, by the homogeneity of the equation $f(X) = 0$, there exists $\mathbf{x} \in \mathbb{Z}^r$ such that $f(\mathbf{x}) = 0$ and $\mathrm{GCD}(x_1, \ldots, x_r) = 1$. Then of course $f(\mathbf{x}) \equiv 0 \pmod{n}$ for every positive integer n. Taking the contrapositive point of view, we note that for each n the existence of such a solution of the congruence $f(X) \equiv 0 \pmod{n}$ is a necessary condition for f to represent 0 in $\mathbb{Q}$. Since these conditions are important, let us say that an integral quadratic form $f(X)$ *represents* 0 *mod* n if and only if there exists $\mathbf{x} \in \mathbb{Z}^r$ such that $f(\mathbf{x}) \equiv 0 \pmod{n}$ and $\mathrm{GCD}(x_1, \ldots, x_r, n) = 1$.

There is another obvious necessary condition for a rational quadratic form f to represent 0 in $\mathbb{Q}$. If all its coefficients a_i are nonzero, then they must not have the same sign, for then (and only then) the form would not even represent 0 in $\mathbb{R}$.

THEOREM 2 (HASSE–MINKOWSKI). *An integral quadratic form f represents zero in $\mathbb{Q}$ if and only if f represents zero mod n for every positive integer n and f represents zero in $\mathbb{R}$.*

A variant of the three-variable case of Theorem 2 was first discovered and proved by Legendre in 1785. The full theorem was formulated and proved in a suitably modern p-adic way by Hasse in 1923 [**7**, pp. 3–22]. Minkowski had earlier (1890) considered the related problem of the classification of rational quadratic forms, finding conditions for a rational quadratic form to represent zero as a consequence. Fine expositions can be found in [**1**] and [**21**].

The Hasse–Minkowski Theorem must be supplemented before it can be used to prove the existence of nontrivial rational solutions of the equation $f(X) = 0$. As it stands, the theorem obliges us to check an infinite number of conditions, one for each n and one more for $\mathbb{R}$. Is that any better than computing $f(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{Q}^r$? What can we do?

To begin with, we need only consider prime power moduli. Suppose that $f(\mathbf{x}_j) \equiv 0 \pmod{n_j}$, $1 \leq j \leq s$, and that $\mathrm{GCD}(n_i, n_j) = 1$ for $i \neq j$. The Chinese Remainder Theorem produces $\mathbf{x} \in \mathbb{Z}^r$ such that $\mathbf{x} \equiv \mathbf{x}_j \pmod{n_j}$ for all j. It follows that $f(\mathbf{x}) \equiv 0 \pmod{\Pi n_j}$. In a typical application the n_j are prime powers. We state the alternative version of Hasse–Minkowski that results.

THEOREM 2′. *An integral quadratic form f represents zero in $\mathbb{Q}$ if and only if f represents zero mod l^n for every prime number l and every $n \geq 1$ and f represents zero in $\mathbb{R}$.*

For each prime number l there is a field $\mathbb{Q}_l$ of characteristic 0, called the field of l-adic numbers, which has the property that an integral quadratic form represents 0 in $\mathbb{Q}_l$ if and only if it represents 0 mod l^n for all $n \geq 1$. Theorem 2′ can therefore be reworded to state that an integral quadratic form represents 0 in $\mathbb{Q}$ if and only if it represents 0 in each of a number of extension fields of $\mathbb{Q}$, which is how Hasse actually stated his theorem. The aesthetic advantage of doing this is that the hypotheses concerning $\mathbb{R}$ and the prime numbers l are put on a uniform footing. The mathematical advantage is that it is much easier to work with the fields $\mathbb{Q}_l$ than with the rings $\mathbb{Z}/l^n\mathbb{Z}$.

We must go further. For an odd prime l the condition that an integral quadratic form f represent zero mod l^n for all n can be expressed in terms of the Legendre symbols (a/l) where the a are divisors of the coefficients of f. Recall that (a/l) is defined for odd primes l and integers a such that $l \nmid a$ by the rule: $(a/l) = +1$ if the congruence $x^2 \equiv a \pmod{l}$ has a solution and $(a/l) = -1$ if the congruence has no solution. We will express the condition that f represent 0 mod 2^n for all n as a set of explicit congruences on the coefficients of f.

LEMMA 3. *Let l be an odd prime, and let a, b, c, d be integers not divisible by l. Then*

(1) *$aX^2 + bY^2$ represents 0 mod l^n for every n if and only if $(-ab/l) = 1$.*

(2) *$aX^2 + bY^2 + cZ^2$ represents 0 mod l^n for every n.*

(3) *$aX^2 + bY^2 + lcZ^2 + ldW^2$ represents 0 mod l^n for every n if and only if $(-ab/l) = 1$ or $(-cd/l) = 1$.*

Proof. We recall the fact (see, e.g., [**6**, p. 67]) that the congruence $x^2 \equiv t \pmod{l^n}$ with t an integer not divisible by the odd prime l has a solution for every n iff $(t/l) = 1$. It follows that if s and t are prime to l and $sx^2 \equiv t \pmod{l}$ has a solution, then $sx^2 \equiv t \pmod{l^n}$ has a solution for every n.

(1). If $(-ab/l) = -1$ then $aX^2 + bY^2$ does not represent 0 mod l. If, on the other hand, $(-ab/l) = +1$, then the congruence $aX^2 \equiv -b$ has a solution mod l (let X be ξa^{-1} where $\xi^2 \equiv -ab \pmod{l}$), whence the congruence has a solution mod l^n, which yields the desired representation of 0 mod l^n (with $Y = 1$).

(2). Let $f = aX^2 + bY^2 + cZ^2$, and let $N = |\{\mathbf{x} \in (\mathbb{Z}/l\mathbb{Z})^3 : f(\mathbf{x}) \equiv 0 \pmod{l}\}|$. We first prove that f represents 0 mod l by showing that $N > 1$.

Fermat's Little Theorem asserts that for $m \in \mathbb{Z}$, $m^{l-1} \equiv 1 \pmod{l}$ if $l \nmid m$, and $m^{l-1} \equiv 0 \pmod{l}$ if $l | m$. Therefore $N \equiv \Sigma\{1 - f(\mathbf{x})^{l-1} : \mathbf{x} \in (\mathbb{Z}/l\mathbb{Z})^3\} \pmod{l}$. Since $1 - f(\mathbf{x})^{l-1}$ has degree $2(l-1)$, it is a linear combination of monomials $X^pY^qZ^r$ where at least one of the exponents p, q, r is less than $l - 1$. But $\Sigma_{w=0}^{l-1} w^s \equiv 0 \pmod{l}$ for $0 \leq s < l - 1$. (If $1 \leq s < l - 1$, write $\Sigma_{w=0}^{l-1} w^s \equiv \Sigma_{k=0}^{l-2} y^{sk} = (1 - y^{s(l-1)})/(1 - y^s) \pmod{l}$, where y is a primitive root mod l, and note that $y^{s(l-1)} \equiv 1$, $y^s \not\equiv 1 \pmod{l}$.) Therefore we can conclude that $N \equiv 0 \pmod{l}$. But $N \neq 0$ because $f(\mathbf{0}) = 0$. It follows that $N \geq l > 1$.

Thus there exists $(x, y, z) \in \mathbb{Z}^3$ such that $f(x, y, z) \equiv 0 \pmod{l}$ and at least one of x, y, z is not divisible by l. If, for example, $l \nmid x$, then the solvability of the congruence $aX^2 = -by^2 - cz^2 \pmod{l}$ (with $X = x$) implies solvability mod l^n for all n. Thus there exists ξ such that $a\xi^2 \equiv -by^2 - cz^2 \pmod{l^n}$ and $X = \xi$, $Y = y$, $Z = z$ gives a representation of 0 mod l^n by f.

(3). If $(-ab/l) = 1$ or $(-cd/l) = 1$, then by (1) f represents 0 mod l^n with $Z = W = 0$ or $X = Y = 0$.

Conversely, suppose that f represents 0 mod l^2 and that $f(x, y, z, w) \equiv 0 \pmod{l^2}$ with x, y, z, w not all divisible by l. If $(-ab/l) = -1$, then the congruence $ax^2 + by^2 \equiv 0 \pmod{l}$ implies that $l|x$ and $l|y$. Therefore $cz^2 + dw^2 \equiv 0 \pmod{l}$. Since l does not divide both z and w, it follows that $(-cd/l) = +1$. □

LEMMA 4. *Let a, b, c, d be odd integers. Then*

(1) $aX^2 + bY^2 + cZ^2$ *represents* 0 *mod* 2^n *for every* n *except when* $a \equiv b \equiv c \pmod{4}$.

(2) $aX^2 + bY^2 + cZ^2 + dW^2$ *represents* 0 *mod* 2^n *for every* n *except when* $a \equiv b \equiv c \equiv d \pmod{4}$ *and* $a + b + c + d \equiv 4 \pmod{8}$.

(3) $aX^2 + bY^2 + cZ^2 + 2dW^2$ *represents* 0 *mod* 2^n *for every* n.

Proof. We recall the fact [**6**, p. 67] that the congruence $X^2 \equiv t \pmod{2^n}$ with t an odd integer has a solution for every n iff $t \equiv 1 \pmod{8}$. It follows that if $sX^2 \equiv t \pmod{8}$ has a solution with s and t odd, then $sX^2 \equiv t \pmod{2^n}$ has a solution for every n.

(1). If $a \equiv b \equiv c \pmod{4}$, then $aX^2 + bY^2 + cZ^2$ does not represent 0 mod 4. On the other hand, suppose, for instance, that $a \not\equiv b \pmod{4}$. Then $a + b \equiv 0$ or $4 \pmod{8}$. Thus $aX^2 + bY^2 + cZ^2 \equiv 0 \pmod{8}$ with $X = Y = 1$ and $Z = 0$ or 2. Letting X satisfy $aX^2 \equiv -b - cZ^2 \pmod{2^n}$ gives a representation of 0 mod 2^n by our form.

(2). If $a \equiv b \equiv c \equiv d \pmod{4}$ and $a + b + c + d \equiv 4 \pmod{8}$ then $aX^2 + bY^2 + cZ^2 + dW^2$ represents 0 mod 4 only by taking X, Y, Z, W all odd and hence does not represent 0 mod 8. If the coefficients are congruent mod 4 and their sum is 0 mod 8, then $X = Y = Z = W = 1$ gives a representation of 0 mod 8 that can be refined as in the proof of (1) to a representation of 0 mod 2^n for every n. Finally, if, for instance, $a \not\equiv b \pmod{4}$ then it follows from (1) that there is a representation of 0 mod 2^n by $aX^2 + bY^2 + cZ^2 + dW^2$ with $W = 0$ for every n.

(3). In view of (1) we may as well assume that $a \equiv b \pmod{4}$. Hence $a + b + 2d \equiv 0$ or $4 \pmod{8}$. It follows that $a + b + cy^2 + 2d \equiv 0 \pmod{8}$ with $y = 0$ or 2. Refinement as before leads to representations of 0 mod 2^n for every n. □

Starting with Lemmas 3 and 4, it is not difficult to show that every integral quadratic form in 5 variables represents 0 mod l^n for every prime l and every $n \geq 1$. Thus the l-adic hypotheses of Theorem 2′ are automatically satisfied by all forms in 5 or more variables, which dramatically reduces arithmetic interest in their coefficients.

We can at last proceed to the number theory in Julia Robinson's definability theorem. The key statements are Lemmas 5 and 6, which is where the Hasse–Minkowski Theorem makes its appearance.

LEMMA 5. *Let $p \equiv 3 \pmod{4}$ be prime, $k \in \mathbb{Z}$, $p \nmid k$, and $4 \nmid k$.*

(1) $f_1 = X^2 + Y^2 - pZ^2 - kW^2$ *represents zero in* $\mathbb{Q}$ *if and only if* $k \not\equiv p \pmod{8}$.

(2) $f_2 = X^2 + Y^2 - pZ^2 - pkW^2$ *represents zero in* $\mathbb{Q}$ *if and only if* $(k/p) = -1$ *and* $pk \not\equiv p \pmod{8}$.

Proof. By Lemma 3(2) applied to $X^2 + Y^2 - pZ^2$, both f_1 and f_2 represent 0 mod l^n for all odd primes $l \neq p$. By Lemma 3(2) applied to $X^2 + Y^2 - kW^2$, f_1 represents 0 mod p^n. By Lemma 3(3) and the fact that $(-1/p) = -1$, f_2 repre-

sents 0 mod p^n iff $(-k/p) = -(k/p) = 1$. If k is even then both f_1 and f_2 represent 0 mod 2^n, by Lemma 4(3). Finally, if k is odd, write $f_i = X^2 + Y^2 - pZ^2 - sW^2$. By Lemma 4(2), f_i represents 0 mod 2^n except in the case where $1 \equiv -p \equiv -s \pmod 4$ and $2 - p - s \equiv 4 \pmod 8$, conditions that are equivalent to the single congruence $s \equiv p \pmod 8$. Thus Lemma 5 follows from Theorem 2'. □

LEMMA 6. *Let p and q be distinct odd primes with $p \equiv 1 \pmod 4$ and $(q/p) = -1$. Let $k \in \mathbb{Z}$ be such that $p \nmid k$, $q \nmid k$, and $4 \nmid k$.*

(1) $f_1 = X^2 + qY^2 - pZ^2 - kW^2$ *represents* 0 *in* $\mathbb{Q}$.

(2) $f_2 = X^2 + qY^2 - pZ^2 - pkW^2$ *represents* 0 *in* $\mathbb{Q}$ *if and only if* $(k/p) = 1$.

(3) $f_3 = X^2 + qY^2 - pZ^2 - qkW^2$ *represents* 0 *in* $\mathbb{Q}$ *if and only if* $(k/q) = 1$.

(4) $f_4 = X^2 + qY^2 - pZ^2 - pqkW^2$ *represents* 0 *in* $\mathbb{Q}$ *if and only if* $(qk/p) = 1$ *and* $(pk/q) = 1$.

Proof. Similar in spirit to the proof of Lemma 5. Consider it an exercise. □

LEMMA 7. *Let n and d be integers with* $\mathrm{GCD}(n, d) = 1$. *Let $p \equiv 3 \pmod 4$ be prime. Set $pn^2 + 2d^2 = st^2$, where $s, t \in \mathbb{Z}$ and s is squarefree. The following two conditions are equivalent.*

(1) *d is odd and not divisible by p.*

(2) *$s \not\equiv p \pmod 8$ and if $p|s$ with $s = pk$, then $(k/p) = -1$.*

Proof. (1) ⇒ (2). If d is odd, then $st^2 \equiv 2 - n^2 \equiv 1$ or 2 (mod 4). But $pt^2 \equiv 0$ or 3 (mod 4). Hence $s \not\equiv p \pmod 8$. If $p \nmid d$, then $p \nmid st^2$, from which it follows that $p \nmid s$.

(2) ⇒ (1). If d is even, then n is odd so $st^2 \equiv p \pmod 8$, from which it follows that t is odd and $s \equiv p \pmod 8$. If $p|d$, then st^2 is divisible by p but not by p^2. Therefore $p|s$ and $p \nmid t$. Writing $s = pk$, we have $kt^2 \equiv n^2 \pmod p$, and therefore $(k/p) = 1$. □

LEMMA 8. *Let n and d be integers with* $\mathrm{GCD}(n, d) = 1$. *Let p and q be odd primes such that $p \equiv 1 \pmod 4$ and $(q/p) = -1$. Set $pqn^2 + 2d^2 = st^2$, where $s, t \in \mathbb{Z}$ and s is squarefree. The following two conditions are equivalent.*

(1) *d is relatively prime to p and q.*

(2) *If $p|s$ with $s = pa$, then $(a/p) = 1$; and if $q|s$ with $s = qb$, then $(b/q) = 1$.*

Proof. (1) ⇒ (2). If d is prime to p and q, then neither p nor q divides s.

(2) ⇒ (1). If $p|d$, then st^2 is divisible by p but not by p^2. Therefore $p|s$ and $p \nmid t$. Writing $s = pa$, we have $at^2 \equiv qn^2 \pmod p$, from which it follows that $(a/p) = (q/p) = -1$. The case $q|d$ is similar, since $(q/p) = (p/q)$. □

Proof of Lemma 1. Let $x \in \mathbb{Q}$, and let $p \equiv 3 \pmod 4$ be prime. Express x in lowest terms as n/d, and let $pn^2 + 2d^2 = st^2$, where $s, t \in \mathbb{Z}$ and s is squarefree. It follows immediately from Lemmas 5 and 7 that $X^2 + Y^2 - pZ^2 - sW^2$ represents 0 in $\mathbb{Q}$ iff d is odd and not divisible by p; and in such a representation of zero we will have $W \neq 0$, since $X^2 + Y^2 - pZ^2$ does not represent 0 (by Lemma 4(1)). Taking $A = tX/dW$, $B = tY/dW$, and $C = tZ/dW$, we find $A^2 + B^2 - pC^2 = px^2 + 2$. □

Proof of Lemma 2. The proof is similar to that of Lemma 1, using Lemmas 6 and 8 in place of Lemmas 5 and 7. To show that $W \neq 0$, it is enough to prove that $f = X^2 + qY^2 - pZ^2$ does not represent 0 mod p^2. We need only show that if $f(x, y, z) \equiv 0 \pmod{p^2}$, then x, y, z are all multiples of p. Since $x^2 + qy^2 \equiv 0$

(mod p) and $(q/p) = -1$, it follows that x and y are divisible by p, and hence that z is also divisible by p. □

As a final application of the Hasse–Minkowski Theorem we prove the rational three-squares theorem cited earlier (Theorem 1). When is a positive integer m a sum of three rational squares? Write $m = 4^a s$ where $a, s \in \mathbb{N}$ and $4 \nmid s$. Then m is a sum of three rational squares if and only if $f = X^2 + Y^2 + Z^2 - sW^2$ represents 0 in $\mathbb{Q}$ (since we cannot have $W = 0$ in such a representation). By Lemma 3(2), f represents 0 mod l^n for all odd primes l and all n (with $W = 0$). By Lemmas 4(2) and 4(3), f represents 0 mod 2^n for all n except in the case $s \equiv 7$ (mod 8). Therefore, by Theorem 2′, m is a sum of three rational squares if and only if $s \not\equiv 7$ (mod 8), which is what we set out to prove.

We close §3 with a simple but important algebraic lemma crucial in many applications of the Hasse–Minkowski Theorem. Though it is avoided in the proof of Robinson's theorem, it could have been used and should not be forgotten.

LEMMA 9. *Let K be a field of characteristic different from 2. Let $f(X) = \sum_{i=1}^r a_i X_i^2$ and $g(X) = \sum_{i=1}^r a_i X_i^2 - sX_{r+1}^2$, where s and all the a_i are nonzero elements of K. Then there exists $\mathbf{x} \in K^r$ such that $f(\mathbf{x}) = s$ if and only if g represents 0 in K.*

Proof. If there exists $\mathbf{x} \in K^r$ such that $f(\mathbf{x}) = s$ then $g(\mathbf{x}, 1) = 0$, showing that g represents 0 in K.

Now suppose that g represents 0 in K. There are two cases to consider.

Case 1. f does not represent 0 in K. If $g(\mathbf{x}) = 0$ with $\mathbf{x} \neq \mathbf{0}$, then $x_{r+1} \neq 0$ and $f(x_1/x_{r+1}, \ldots, x_r/x_{r+1}) = s$.

Case 2. f does represent 0 in K. Let $\mathbf{x} \neq \mathbf{0} \in K^r$ satisfy $f(\mathbf{x}) = 0$. Find $\mathbf{y} \in K^r$ such that $f(\mathbf{y}) = 0$ and $\sum_{i=1}^r a_i x_i y_i = 1$. Then $f(\mathbf{x} + \frac{1}{2}s\mathbf{y}) = s$. Produce $\mathbf{y}$ as follows. Choose $\mathbf{u} \in K^r$ such that $\sum_{i=1}^r a_i x_i u_i = 1$. Take $\mathbf{y} = \mathbf{u} + \lambda \mathbf{x}$, where $\lambda = -f(\mathbf{u})/2$. □

4. Cubic forms. From the point of view of number theory, the Hasse–Minkowski Theorem used in §3 is rather special. Straightforward generalizations of Theorem 2 to other classes of Diophantine equations are usually false. Such is the case even for homogeneous cubic equations, seemingly the simplest equations beyond those covered by Theorem 2.

It is easy nowadays, using computers, to find likely candidates for counterexamples to a Generalized Hasse–Minkowski Theorem, but it is not so easy to prove that they are true counterexamples. One has got to show that a specific Diophantine equation has no rational solutions, and congruence considerations will not help. The method has usually been reliance on ad hoc computations.

The theory of integral cubic forms[1] in 3 variables is inextricably bound up with the theory of elliptic curves. Selmer (1951) was the first to find a 3-variable cubic counterexample to Hasse–Minkowski: $3X^3 + 4Y^3 + 5Z^3$ represents 0 mod n for every n but fails to represent 0 in $\mathbb{Q}$.

It was thought for several years that Hasse–Minkowski might hold for homogeneous cubic equations in 4 variables. Counterexamples proved very hard to construct, but they were found, the earliest by Swinnerton-Dyer (1962) and Cassels

[1] A *form* is simply a homogeneous polynomial.

and Guy (1966). Recently Colliot-Thélène, Kanevsky, and Sansuc (1985) found the following nontrivial infinite family of counterexamples: $X^3 + p^2Y^3 + pqZ^3 + q^2W^3$, where p and q are prime numbers, $p \equiv 2$, $q \equiv 5 \pmod 9$, represents 0 mod n for all n but fails to represent 0 in $\mathbb{Q}$. An idea of Manin (1970) provides a uniform explanation for all currently known 4-variable cubic counterexamples to Hasse–Minkowski.

It is possible that the obvious analogue of the Hasse–Minkowski Theorem is true for cubic forms in 5 or more variables. No counterexamples have been found. There can't be any counterexamples in 16 or more variables, since *every* integral cubic form in 16 or more variables represents 0 in $\mathbb{Q}$, by a theorem of Davenport (1963). The conjecture that every integral cubic form in 10 or more variables represents zero in $\mathbb{Q}$ has been proved for nonsingular forms by Heath-Brown (1983).

Though there has been recent progress, it remains an open problem to find a usable set of necessary and sufficient conditions that a cubic form represent 0 in $\mathbb{Q}$. The *Hasse principle*, the working hypothesis that there should be some kind of strong relation between local (mod n) solvability and global ($\mathbb{Q}$) solvability, seems to be a reliable guide, even if not precise.

See [**11**] and [**19**] for a discussion of these and related results.

5. Hilbert's Tenth Problem. The definition of $\mathbb{Z}$ in $\mathbb{Q}$ that Julia Robinson discovered is complicated, especially when compared to the definition of $\mathbb{N}$ in $\mathbb{Z}$ using the four-squares theorem, where only existential quantifiers are used. An interesting open question is whether such a simple definition exists for defining $\mathbb{Z}$ in $\mathbb{Q}$ (i.e., Is $\mathbb{Z}$ Diophantine over $\mathbb{Q}$?). This problem is especially important because of its connection with Hilbert's Tenth Problem, which asked for an algorithm to decide whether an arbitrary Diophantine equation has a solution in $\mathbb{Z}$. Hilbert's problem was solved, negatively, in 1970, by Y. Matijasevič who completed the final step in a program developed by Julia Robinson, Martin Davis, and Hilary Putnam (see [**2**]). In fact, Matijasevič has shown (see [**8**]) that there is no algorithm for solving 9-variable Diophantine equations over $\mathbb{N}$. Because $\mathbb{N}$ is Diophantine over $\mathbb{Z}$, this implies that there is no algorithm for solving Diophantine equations over $\mathbb{Z}$. For such an algorithm over $\mathbb{Z}$ would yield one over $\mathbb{N}$, as we illustrate in the case of two variables: $P(X, Y) = 0$ has a solution in $\mathbb{N}^2$ if and only if

$$P(A^2 + B^2 + C^2 + D^2, E^2 + F^2 + G^2 + H^2) = 0$$

has a solution in $\mathbb{Z}^8$. Because $4 \cdot 9 = 36$, this technique shows that there is no algorithm for solving 36-variable equations over $\mathbb{Z}$. We digress momentarily to show how 36 can be reduced to 27, a consequence of the following 3-variable definition of $\mathbb{N}$ in $\mathbb{Z}$: An integer X is nonnegative iff X has the form $A^2 + B^2 + C^2 + C$. The validity of this definition (found by Raphael Robinson) can be seen by using the integer three-squares theorem (§2) to write $4X + 1$ as a sum of three squares, of which two must be even and one odd. Alternatively, one could use the fact that every nonnegative integer is a sum of three triangular numbers, also a consequence of the three-squares theorem (see [**6**, p. 179]). Raphael Robinson [**16**] found a two-variable definition of $\mathbb{N}$ in $\mathbb{Z}$ by using properties of the Pell equation, but X is not expressed as a polynomial in the two new variables, and the definition also includes an inequation, so it does not serve to further reduce the 27 to 18. However, Matijasevič recently modified this idea, finding that an integer A is nonnegative iff there are integers X and Y with $Y \neq 0$ such that $X^2 -$

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A Simple Proof of a Simple Version of the Riemann Mapping Theorem by Simple Functional Analysis

P. R. GARABEDIAN
Department of Mathematics, New York University, NY, NY 10012

The purpose of this note is to sketch a proof of a restricted version of the Riemann mapping theorem that only requires an elementary knowledge of complex analysis and avoids concepts like completeness or compactness. The proof is motivated by the relationship of the map function to the Green's function and the Szegö kernel function [1].

Let D be a simply connected region of the complex plane bounded by a smooth Jordan curve ∂D enclosing the origin, and let A be the class of functions $f(z)$ analytic in the closure of D. Consider the problem of minimizing the norm

$$\|f\|^2 = \int_{\partial D} |f|^2 \, |dz|$$

subject to the constraint

$$f(0) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(z)}{z} \, dz = 1,$$

which ensures that $\|f\|^2$ has an infimum $d > 0$ by virtue of the inequality

$$\left| \int_{\partial D} \frac{f(z)}{z} \, dz \right|^2 \leq \left(\max_{\partial D} \frac{1}{|z|^2} \right) \int_{\partial D} |dz| \int_{\partial D} |f|^2 |dz| .$$

Because $(f_m + f_n)/2$ is a normalized function in A whenever f_m and f_n are, so that $\|f_m + f_n\|^2 \geq 4d$, the parallelogram law

$$\|f_m - f_n\|^2 + \|f_m + f_n\|^2 = 2\|f_m\|^2 + 2\|f_n\|^2$$

implies that any minimal sequence f_n for this extremal problem is a Cauchy sequence with $\|f_m - f_n\| \to 0$. Moreover, because $\|f_n\|$ remains bounded, we see from the estimate

$$\begin{aligned} |(gf_m, f_m) - (gf_n, f_n)| &\leq |(gf_m - gf_n, f_m)| + |(gf_n, f_m - f_n)| \\ &\leq \left(\max_{\partial D} |g| \right) (\|f_m\| + \|f_n\|) \|f_m - f_n\| \end{aligned}$$

that the sequence of complex numbers (gf_n, f_n) converges and the formula

$$L[g] = \lim_{n \to \infty} \frac{(gf_n, f_n)}{d} = \lim_{n \to \infty} \frac{1}{d} \int_{\partial D} g |f_n|^2 |dz|$$

defines a bounded linear functional over the space of continuous functions g on the curve ∂D.

If ε is any complex number and g is in A with $g(0) = 0$, then $f_n + \varepsilon g$ is a normalized element of A and, therefore,

$$d \leq \|f_n + \varepsilon g\|^2 = \|f_n\|^2 + 2\,\mathrm{Re}\{\varepsilon(gf_n, f_n)\} + |\varepsilon|^2\|gf_n\|^2.$$

Since $|g|$ is bounded on ∂D there is a fixed number K such that $\|gf_n\|^2 \leq K$, so that letting $n \to \infty$ we have

$$d \leq d + 2\,\mathrm{Re}\{\varepsilon L[g]\} + K|\varepsilon|^2.$$

That is only possible for arbitrary ε if $L[g] = 0$. Since $L[1] = 1$, substitution of the renormalized function $g(z) - g(0)$ into this equation yields the variational condition

$$g(0) = L[g].$$

Hence when w^* is any point outside D we can put $g(z) = \log(z - w^*)$ to arrive at the identity

$$\log|w^*| - L[\log|z - w^*|] = 0.$$

For $w = u + iv$ inside D we proceed to establish that the formula

$$\log|F(w)| = \log|w| - L[\log|z - w|]$$

defines by analytic completion a function $\zeta = F(w)$ mapping D conformally onto the interior of the unit circle $|\zeta| < 1$.

If $g(z, t)$ depends on a parameter t in such a way that

$$\lim_{t \to t_0} \left| \frac{g(z,t) - g(z,t_0)}{t - t_0} - \frac{\partial g}{\partial t} \right| = 0$$

uniformly for all z on ∂D, then the boundedness of the linear functional L implies that

$$\frac{\partial}{\partial t} L[g] = L\left[\frac{\partial g}{\partial t}\right].$$

Consequently L commutes with the Laplace operator

$$\Delta = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}$$

in the special case $g(z) = \log|z - w|$, which shows that $\log|F|$ is harmonic,

$$\Delta \log|F| = \Delta \log|w| - L[\Delta \log|z - w|] = 0.$$

The corresponding analytic function $F(w)$ is seen by inspection to have a simple zero at the origin.

If w and w^* are situated at equal distances from ∂D along the same normal, then

$$\lim_{|w - w^*| \to 0} \log\left|\frac{z - w}{z - w^*}\right| = 0$$

uniformly for all z on ∂D. The reason for this is that for any $\delta > 0$ the region of

the z-plane satisfying

$$\frac{1}{1+\delta} < \left|\frac{z-w}{z-w^*}\right| < 1+\delta$$

is bounded by two disjoint circles of a coaxial system; and because ∂D is smooth, it will lie entirely inside any such region when w and w^* are sufficiently close. It follows that

$$\lim \log|F| = \lim \log\frac{|w|}{|w^*|} - \lim L\left[\log\left|\frac{z-w}{z-w^*}\right|\right] = 0$$

as w and w^* approach the same point on the boundary ∂D, so $|F| = 1$ there. From the maximum principle we conclude that $|F| < 1$, and from the argument principle we conclude that F is univalent, since it has only one simple zero. Thus it is the desired map function, and we have proved the

THEOREM. *Any simply connected region of the plane bounded by a smooth Jordan curve can be mapped conformally onto the unit circle.*

This work was partially supported by NSF Grant DMS-8922805.

REFERENCE

1. P. R. Garabedian, Partial Differential Equations, Chelsea, 1986.

On the Equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$

ALFONSO VILLANI
Department of Mathematics, University of Messina, Messina, Italy

The starting point of this note is the consideration of two quite obvious facts. First, if we are given a Lebesgue measurable set $E \subseteq \mathbb{R}^n$ and a function $u: E \to \mathbb{R}$ which is locally Lebesgue integrable on E, that is integrable on every compact subset of E, then u is not integrable over all E in general. Second, there are, however, Lebesgue measurable sets E for which local integrability of a function implies integrability over all E; indeed, all sets E which are "essentially compact", that is which contain a compact set H such that $E \setminus H$ has measure zero, have this property.

Thus, a natural question is: for which E's does local integrability imply integrability? In a more abstract setting, one can ask about a characterization of those measure spaces for which every locally integrable (in a sense to be specified) function is integrable on the entire space. Providing such a characterization is the aim of this note. It turns out, in particular, (see the Corollary below) that the answer to the starting question is: exactly those E's which are essentially compact.

Let $(\Omega, \mathscr{A}, \mu)$ be a measure space. We will assume throughout that

i) Ω *is a metrizable space*; and

ii) *if* $K \subseteq \Omega$ *is a compact set, then* $K \in \mathscr{A}$ *and* $\mu(K) < \infty$.

A function $u: \Omega \to \mathbb{R}$ is said to be *locally* μ*-integrable* provided that u is $\mathscr{A}$-measurable and $\int_K |u|\, d\mu < \infty$ for every compact set $K \subseteq \Omega$. The set of all

the z-plane satisfying

$$\frac{1}{1+\delta} < \left|\frac{z-w}{z-w^*}\right| < 1+\delta$$

is bounded by two disjoint circles of a coaxial system; and because ∂D is smooth, it will lie entirely inside any such region when w and w^* are sufficiently close. It follows that

$$\lim \log|F| = \lim \log\frac{|w|}{|w^*|} - \lim L\left[\log\left|\frac{z-w}{z-w^*}\right|\right] = 0$$

as w and w^* approach the same point on the boundary ∂D, so $|F| = 1$ there. From the maximum principle we conclude that $|F| < 1$, and from the argument principle we conclude that F is univalent, since it has only one simple zero. Thus it is the desired map function, and we have proved the

THEOREM. *Any simply connected region of the plane bounded by a smooth Jordan curve can be mapped conformally onto the unit circle.*

This work was partially supported by NSF Grant DMS-8922805.

REFERENCE

1. P. R. Garabedian, Partial Differential Equations, Chelsea, 1986.

On the Equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$

ALFONSO VILLANI
Department of Mathematics, University of Messina, Messina, Italy

The starting point of this note is the consideration of two quite obvious facts. First, if we are given a Lebesgue measurable set $E \subseteq \mathbb{R}^n$ and a function $u: E \to \mathbb{R}$ which is locally Lebesgue integrable on E, that is integrable on every compact subset of E, then u is not integrable over all E in general. Second, there are, however, Lebesgue measurable sets E for which local integrability of a function implies integrability over all E; indeed, all sets E which are "essentially compact", that is which contain a compact set H such that $E \setminus H$ has measure zero, have this property.

Thus, a natural question is: for which E's does local integrability imply integrability? In a more abstract setting, one can ask about a characterization of those measure spaces for which every locally integrable (in a sense to be specified) function is integrable on the entire space. Providing such a characterization is the aim of this note. It turns out, in particular, (see the Corollary below) that the answer to the starting question is: exactly those E's which are essentially compact.

Let $(\Omega, \mathscr{A}, \mu)$ be a measure space. We will assume throughout that

i) Ω *is a metrizable space*; and

ii) *if* $K \subseteq \Omega$ *is a compact set, then* $K \in \mathscr{A}$ *and* $\mu(K) < \infty$.

A function $u: \Omega \to \mathbb{R}$ is said to be *locally* μ*-integrable* provided that u is $\mathscr{A}$-measurable and $\int_K |u|\, d\mu < \infty$ for every compact set $K \subseteq \Omega$. The set of all

locally μ-integrable functions is denoted by $L^1_{\text{loc}}(\mu)$. Of course we have $L^1(\mu) \subseteq L^1_{\text{loc}}(\mu)$.

It is worth recalling here before stating the Theorem below that a set $A \in \mathscr{A}$, with $\mu(A) > 0$, is said to be a *μ-atom* provided that if $B \in \mathscr{A}$, $B \subseteq A$, then either $\mu(B) = 0$ or $\mu(A \setminus B) = 0$.

THEOREM. *The following condition* $(*)$ *is necessary and sufficient in order that* $L^1(\mu) = L^1_{\text{loc}}(\mu)$:

$(*)$ *there exists a compact set* $H \subseteq \Omega$ *such that either* $\mu(\Omega \setminus H) = 0$ *or otherwise* $\Omega \setminus H$ *is a finite union of* μ*-atoms of finite measure.*

We need the following easy topological lemma.

LEMMA. *Let* K *be a compact subset of a Hausdorff space* Ω *and let* $G_1, \ldots, G_r$ *be open subsets of* Ω *such that* $K \subseteq G_1 \cup \cdots \cup G_r$. *Then there exist* $K_1, \ldots, K_r$, *compact subsets of* Ω, *with* $K_i \subseteq G_i$, $i = 1, \ldots, r$, *such that* $K = K_1 \cup \cdots \cup K_r$.

Proof. Consider first the case $r = 2$.

Let $H_i = K \cap (\partial G_i)$, $i = 1, 2$. Then H_1, H_2 are disjoint compact sets, with $H_1 \subseteq G_2$, $H_2 \subseteq G_1$. Since Ω is a Hausdorff space, there exist A_1, A_2, disjoint open sets, with $H_i \subseteq A_i$, $i = 1, 2$. Obviously we can assume $A_1 \subseteq G_2$, $A_2 \subseteq G_1$. Let $B_1 = A_1 \cup (G_2 \setminus \overline{G}_1)$, $B_2 = A_2 \cup (G_1 \setminus \overline{G}_2)$. Then B_1, B_2 are disjoint open sets, with $K \setminus G_i \subseteq B_i$, $i = 1, 2$. Consequently, if we put $K_i = K \setminus B_i$, $i = 1, 2$, then K_1, K_2 are compact sets, with $K = K_1 \cup K_2$ and $K_i \subseteq G_i$, $i = 1, 2$.

An easy inductive argument completes the proof.

Proof of the Theorem. It is well known (see [1], Lemma 6.5.8) that if $u\colon \Omega \to \mathbb{R}$ is $\mathscr{A}$measurable and $A \in \mathscr{A}$ is a μ-atom, then u is μ-almost everywhere constant on A; hence we have $\int_A |u|\, d\mu < \infty$ provided that $\mu(A) < \infty$. Thus, it is clear that if condition $(*)$ is satisfied, then $L^1(\mu) = L^1_{\text{loc}}(\mu)$.

To prove the reverse implication, assume the equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$ and consider the collection $\mathscr{V}$ of all open sets $G \subseteq \Omega$ containing no compact set of positive measure. Let $V = \cup\{G\colon G \in \mathscr{V}\}$. In other words, V is the set of all points $x \in \Omega$ which have the following property: there exists an open neighbourhood of x which contains no compact set of positive measure. We will prove that:

a) $V \in \mathscr{V}$,

b) $H = \Omega \setminus V$ *is a compact set*,

c) *if* $\mu(V) > 0$, *then* V *is a finite union of* μ*-atoms of finite measure.*

Of course, b) and c) imply condition $(*)$.

To prove a), consider any compact set $K \subseteq V$. By the definition of V and the compactness of K there exist finitely many sets $G_1, \ldots, G_r \in \mathscr{V}$ such that $K \subseteq G_1 \cup \cdots \cup G_r$. According to the Lemma, there are compact sets $K_1, \ldots, K_r$, with $K_i \subseteq G_i$ (and hence $\mu(K_i) = 0$), $i = 1, \ldots, r$, such that $K = K_1 \cup \cdots \cup K_r$. It follows that $\mu(K) = 0$ and this proves assertion a).

Next, to prove b), assume by contradiction that there exists a sequence $\{x_n\}$ in H such that no subsequence is convergent in H, and hence in Ω since H is closed. Of course, it can be also assumed that $x_n \neq x_m$ if $n \neq m$. Then it is easy to construct a sequence $\{B_n\}$ of pairwise disjoint open subsets of Ω, with $x_n \in B_n$ for every $n = 1, 2, \ldots$. Moreover, it can be assumed that the diameter of B_n, with respect to any fixed compatible metric d on Ω, tends to zero as $n \to \infty$. As a

The Ranks of Tournament Matrices

D. DE CAEN
Department of Mathematics and Statistics, Queen's University, Kingston, Canada K7L 3N6

A *tournament* matrix of order n is an $n \times n$ matrix M of 0's and 1's such that $M_{ii} = 0$ and $M_{ij} + M_{ji} = 1$ for all $i \neq j$. In other words, M satisfies the matrix equation

$$M + M^t = J - I, \tag{1}$$

where M^t is the transpose of M, J is the $n \times n$ matrix of all 1's, and I is the identity matrix. See [**1**] for background information on tournaments. Since M has only $\{0, 1\}$-entries, it may be viewed as a matrix over any field F. We shall make a few observations concerning the rank of M over F.

First of all, we mention a simple definition of the rank of any $m \times n$ matrix A over F. This interpretation of rank is perhaps not widely known, but is occasionally quite useful. A *rank-one* matrix is, by definition, a matrix of the form uv^t, where u is a non-zero $m \times 1$ vector and v is a non-zero $n \times 1$ vector. The rank of a non-zero matrix A is then defined as the *minimum number of rank-one matrices that sum* to A. (By convention, the zero matrix has rank 0.) We leave as an exercise the verification that this definition of rank is the same as the usual one, namely the dimension of the row space of A. Incidentally, our definition of rank makes it quite obvious that the row and column ranks of A are equal. This viewpoint originated in multilinear algebra [**6**, p. 38] when defining the rank of symmetric tensors.

THEOREM 1. *Let M be an $n \times n$ tournament matrix. Then the rank of M over any field is at least $(n - 1)/2$.*

Proof. The rank of $J - I$ is always at least $n - 1$. Indeed, $\det(J - I)$ is $\pm(n - 1)$, but it also contains the $(J - I)$-matrix of order $(n - 1)$ as a principal submatrix, which has determinant $\pm(n - 2)$. Thus, even if the characteristic of F divides $(n - 1)$, $J - I$ still has a non-vanishing minor of order $(n - 1)$; so $\operatorname{rank}(J - I) \geqslant n - 1$. The theorem now follows from equation (1). Indeed, if M were expressible as a sum of fewer than $(n - 1)/2$ rank-one matrices, then so would M^t, and so by (1) $J - I$ would have rank less than $(n - 1)$, a contradiction. ∎

We say that a matrix M is *regular* if it has constant row and column sums. Thus M is regular if and only if $MJ = JM = kJ$, for some k. If M is a regular $n \times n$ tournament matrix, then clearly $k = (n - 1)/2$, and n must be odd.

THEOREM 2. *Let M be an $n \times n$ regular tournament matrix. Then the rank of M over any field of characteristic not dividing $(n - 1)/2$ is at least $(n + 1)/2$.*

Proof. Suppose that $\operatorname{rank}(M) = (n - 1)/2 = k$, so that $M = u_1v_1^t + \cdots + u_kv_k^t$ for some rank-one matrices $u_iv_i^t$. Then $u_1, \ldots, u_k$ span the column space of M, and $v_1, \ldots, v_k$ span the column space of M^t. Thus by (1), these $2k$ vectors altogether span the column space of $J - I$, which has dimension $(n - 1) = 2k$ so these vectors are independent. However, since M is regular, the column spaces of

M and M^t have a common 1-dimensional subspace, namely the one spanned by the vector $(1, 1, \ldots, 1)^t$, which is a contradiction. ■

Here is an application of Theorem 2. Let q be a prime-power congruent to 3 modulo 4. The *Paley tournament* matrix $M(q)$ is the $q \times q$ matrix with rows and columns labelled with the elements of the finite field $GF(q)$, with the (i, j)-entry equal to 1 if $i - j$ is a non-zero square in $GF(q)$, and 0 otherwise. Lander [**5**], p. 138, uses a bit of character theory to show that $\operatorname{rank}(M(q)) \geqslant (q + 1)/2$. Since $M(q)$ is a regular tournament matrix, we get a simpler proof using Theorem 2. Lander further notes that if p is a prime divisor of $(q + 1)/4$, then the rank of $M(q)$ over $GF(p)$ is at most $(q + 1)/2$, so Theorem 2 is in general best possible. However, Theorems 1 and 2 can be improved upon if the underlying field has characteristic zero.

THEOREM 3. *Let M be an $n \times n$ tournament matrix. Then the rank of M over any field of characteristic zero is at least $(n - 1)$.*

Proof. It suffices to prove this for the field of real numbers (for then M will have a submatrix of order $n - 1$ of non-zero determinant, and this will be non-zero over any field of characteristic zero). We want to show that the kernel of M has dimension at most 1. If $Mx = 0$ for some vector x, then (1) implies that $x^t(J - I)x = 0$. We claim that the set of vectors satisfying this last equation contains no 2-dimensional subspace, which implies that $\dim(\ker(M)) \leqslant 1$.

Note that $J - I$ has rank n and index 1, so we may work instead with the equivalent quadratic equation (here $x^t = (x_1, x_2, \ldots, x_n)$)

$$x_1^2 = x_2^2 + \cdots + x_n^2. \tag{2}$$

Let S be the set of solutions to (2); we will show that it does not contain a 2-dimensional subspace W. (This is an elementary property of the isotropic cone of Minkowski n-space; see Efimov [**4**], section 7.2, for an introduction to Minkowski space.) Let x and y be basis vectors for W. Then $ax + by \in W \subseteq S$ for all scalars a and b, so $ax + by$ satisfies (2). We easily derive from this the relation

$$x_1 y_1 = x_2 y_2 + \cdots + x_n y_n. \tag{3}$$

Now applying (2), (3) and the Schwarz inequality, we obtain

$$\begin{aligned} x_1^2 y_1^2 &= |x_2 y_2 + \cdots + x_{nn}|^2 \\ &\leqslant \left(x_2^2 + \cdots + x_n^2\right)\left(y_2^2 + \cdots + y_n^2\right) \\ &= x_1^2 y_1^2. \end{aligned}$$

Thus, equality holds in the Schwarz inequality, and so $(y_2, \ldots, y_n) = c(x_2, \ldots, x_n)$, for some scalar c. Using (2) we easily deduce that $y_1^2 = c^2 x_1^2$, so $y_1 = \pm c x_1$. Note that we cannot have $y_1 = -c x_1$: otherwise, using (3) we obtain $-c x_1^2 = c x_1^2$, and so either $c = 0$ (in which case y is the zero vector, which is impossible), or $x_1 = 0$ (and so by (2) x is the zero vector, which is impossible). Thus $y_1 = c x_1$ and so $y = cx$, a final contradiction. ■

An entirely different proof of Theorem 3, due to D. A. Gregory, appears in [**3**]. N. J. Pullman (personal communication) has obtained the following generalization of Theorem 3. If an eigenvalue of a tournament matrix has real part not equal to $-1/2$, then its geometric multiplicity equals one.

THEOREM 4. *Let M be an $n \times n$ regular tournament matrix with $n \geqslant 3$. Then M is nonsingular, provided the underlying field F has characteristic zero.*

Proof. Suppose $Mx = 0$. Then by (1), $0 = x^t(J - I)x = x^tJx - x^tx$. Moreover, $kJx = JMx = 0$, so $x^tJx = 0$ and we get $0 = x^tx = x_1^2 + \cdots + x_n^2$. Since F may be assumed to be the real field, we conclude that x is the zero vector. ■

Theorem 4 was first proved by Brauer and Gentry [2], by a different method. An interesting application of this theorem to a graph decomposition problem is given in [3]. In conclusion, we mention another interesting property of regular tournament matrices: within the class of tournament matrices, they coincide with the normal matrices, i.e. those satisfying $MM^t = M^tM$; cf. [2; thm. 2]. Indeed, since $MM^t = M(J - I - M)$ and $M^tM = (J - I - M)M$, $MM^t = M^tM$ if and only if $MJ = JM$, that is if and only if M is regular.

REFERENCES

1. L. W. Beineke and K. B. Reid, Tournaments, in Selected Topics in Graph Theory (Beineke and Wilson, eds.), Academic Press, 1978.
2. A. Brauer and I. C. Gentry, Some remarks on tournament matrices, *Linear Algebra and Appl.*, 5 (1972) 311–318.
3. D. de Caen and D. G. Hoffman, Impossibility of decomposing the complete graph on n points into $(n - 1)$ isomorphic complete bipartite subgraphs, *SIAM J. on Discrete Math.*, 2 (1989) 48–50.
4. N. V. Efimov, Higher Geometry, Mir Publishers, 1980.
5. E. S. Lander, Symmetric Designs: An Algebraic Approach, Cambridge Univ. Press, 1983.
6. M. Marcus, Finite Dimensional Multilinear Algebra, vol. 1, Dekker, 1973.

Convergence-Preserving Functions: An Alternative Discussion

ARTHUR SMITH
Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5S6, Canada

In a recent article in this publication [1] a discussion was presented on *convergence-preserving* (real) functions f, ones for which Σa_n convergent implies $\Sigma f(a_n)$ convergent. It follows necessarily that $f(x)$ must be a constant multiple of x in some neighbourhood of the origin.

This problem has been around for some time and its discussion in the MONTHLY reminded me of the time I stumped some colleagues by asking for an example of sequence $\{a_n\}$ such that Σa_n converges but $\Sigma \sin a_n$ does not.

Below I give a proof that convergence-preserving (CP) functions must be of the form described above, which seems somewhat shorter than that given previously and does not require considering the differentiability of such functions.

At the end I solve the specific problem posed above which the reader may like to try before she/he reads on.

First, I show f is finitely additive in some neighbourhood of the origin.

Suppose f is CP and let

$$g(x, y) = f(x + y) + f(-x) + f(-y).$$

Assume for the moment there is no neighbourhood of the origin in which $g(x, y) \equiv 0$. Then there are null sequences $\{x_n\}, \{y_n\}$ with $g(x_n, y_n) \neq 0$. Let []

denote "integer part of" and let

$$M_n = \left[\left|g(x_n, y_n)^{-1}\right|\right] + 1,$$

and consider the series S below, whose sum to M terms is S_M:

$$(x_1 + y_1) - x_1 - y_1 + (x_1 + y_1) - x_1 - y_1 + \cdots + (x_1 + y_1) - x_1 - y_1$$
$$+ (x_2 - y_2) - x_2 - y_2 + \cdots + (x_2 + y_2) - x_2 - y_2 + \cdots + (x_n + y_n)$$
$$- x_n - y_n + \cdots,$$

where the group of terms $(x_n + y_n) - x_n - y_n$ is repeated M_n times. The series converges to zero since $S_{3k} = 0, S_{3k+1} = x_n + y_n, S_{3k+2} = y_n$ for some $n = n(k)$ where $n(k) \to \infty$ as $k \to \infty$.

Rewrite S as $\sum a_m$. Then $\sum f(a_m)$ converges, and so by grouping the terms we see $\sum M_n g(x_n, y_n)$ converges. This is impossible since $|M_n g(x_n, y_n)| > 1$ by construction.

Thus $g(x, y) \equiv 0$ in some neighbourhood of the origin, i.e.

$$f(x + y) + f(-x) + f(-y) = 0 \qquad (*)$$

for $|x| + |y|$ sufficiently small.

Put $x = y = 0$ in $(*)$ to show $f(0) = 0$ (which can be proved directly by considering the convergent series all of whose terms are zero). Now put $y = 0$ in $(*)$ to show f is odd so that

$$f(x + y) = f(x) + f(y) \qquad (**)$$

in some neighbourhood N of zero. This condition together with continuity of f at zero is sufficient to show $f(x) = kx$ for $|x|$ sufficiently small and k a constant.

Continuity of f at 0 is established as follows.

Consider any null sequence $\{a_n\}$ and form the convergent series

$$a_1 - a_1 + a_2 - a_2 + \cdots + a_n - a_n + \cdots.$$

Since f is odd it follows that

$$f(a_1) - f(a_1) + f(a_2) - f(a_2) + \cdots + f(a_n) - f(a_n) + \cdots$$

converges, so $f(a_n) \to 0$ as $a_n \to 0$, and hence $f(x) \to 0$ as $x \to 0$ and continuity at 0 is established since $f(0) = 0$.

Continuity in N now follows by letting $x \to -y$ in $(**)$. Let x_0 be a nonzero point of N. It is easy to show $f(x_0/n) = (1/n)f(x_0)$ for any nonzero integer n and hence prove that $f(rx_0) = rf(x_0)$ for any rational r with $|r| \leqslant 1$. The continuity of f now implies that $f(x) = kx$, $x \in [-x_0, x_0]$, where $k = f(x_0)/x_0$.

To find a sequence $\{a_n\}$ such that $\sum a_n$ converges but $\sum \sin a_n$ does not, we use a construction similar to that used in deducing the contradiction in the proof above.

Put $x_n = y_n = 1/n$ in the above proof and note $g(x_n, y_n) = \sin(2/n) - 2\sin(1/n) \neq 0$ since, in fact, $g(x_n, y_n) < 0$ for each n. Define M_n as in the general proof (or, more simply, $M_n = n^3$ since $n^3 g(1/n, 1/n) \to -1 \neq 0$). Then

$$1 + 1 - 2 + \cdots + \frac{1}{2} + \frac{1}{2} - 1 + \cdots$$
$$+ \cdots + \frac{1}{n} + \frac{1}{n} - \frac{2}{n} + \cdots +,$$

where each group $1/n + 1/n - 2/n$ is repeated M_n times, converges while the sum of the sines of the terms does not.

I am obliged to the referee for suggesting an example like the one above which mirrors the general construction more closely than my original example.

REFERENCE

1. Gerald Wildenburg, Convergence-preserving functions, *Amer. Math. Monthly*, 95 (1988) 542–544.

Crossing the Square

KEITH WHITTINGTON
Department of Mathematics, University of the Pacific, Stockton, CA 95211

This note gives a counterexample to the conjecture in [**1**] that there are no "level crossings" of the square, and then concludes by indicating that in two important ways, this example is as simple as possible.

A **crossing** of the square $S = [0, 1] \times [0, 1]$ is a pair $[X, Y]$ of disjoint, connected subsets of S, each containing two opposite corners. The crossing is **level** if $X \cup Y$ does not contain disjoint, connected sets E and F such that each contains two adjacent corners of S.

Our example, depicted in FIGURE 1, consists of the sets

$$X = \{p\} \cup \{p'\} \cup A_1 \cup B_2 \cup A_3 \cup B_4 \cup \cdots$$
$$Y = C \cup B_1 \cup A_2 \cup B_3 \cup A_4 \cup \cdots.$$

The crossing $[X, Y]$ is indeed level, but in fact, there are no two nondegenerate, disjoint, connected sets E and F in $X \cup Y$ such that $p \in E$ and $p' \in F$.

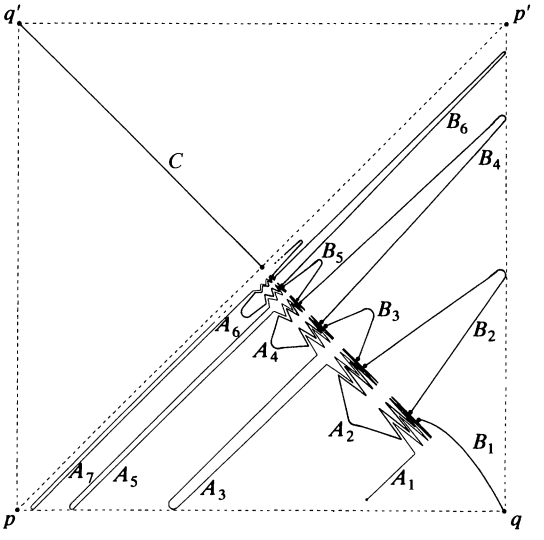


FIG. 1

where each group $1/n + 1/n - 2/n$ is repeated M_n times, converges while the sum of the sines of the terms does not.

I am obliged to the referee for suggesting an example like the one above which mirrors the general construction more closely than my original example.

REFERENCE

1. Gerald Wildenburg, Convergence-preserving functions, *Amer. Math. Monthly*, 95 (1988) 542–544.

Crossing the Square

KEITH WHITTINGTON
Department of Mathematics, University of the Pacific, Stockton, CA 95211

This note gives a counterexample to the conjecture in [**1**] that there are no "level crossings" of the square, and then concludes by indicating that in two important ways, this example is as simple as possible.

A **crossing** of the square $S = [0, 1] \times [0, 1]$ is a pair $[X, Y]$ of disjoint, connected subsets of S, each containing two opposite corners. The crossing is **level** if $X \cup Y$ does not contain disjoint, connected sets E and F such that each contains two adjacent corners of S.

Our example, depicted in FIGURE 1, consists of the sets

$$X = \{p\} \cup \{p'\} \cup A_1 \cup B_2 \cup A_3 \cup B_4 \cup \cdots$$
$$Y = C \cup B_1 \cup A_2 \cup B_3 \cup A_4 \cup \cdots.$$

The crossing $[X, Y]$ is indeed level, but in fact, there are no two nondegenerate, disjoint, connected sets E and F in $X \cup Y$ such that $p \in E$ and $p' \in F$.

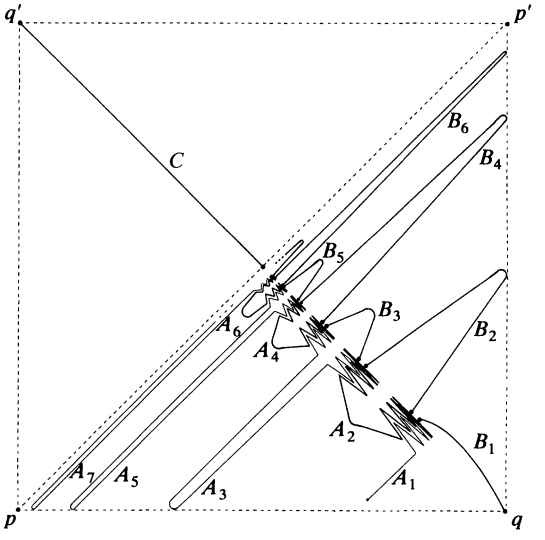


FIG. 1

Suppose such sets E and F exist. Then E must meet A's of arbitrarily large odd subscript, and F must meet B's of arbitrarily large even subscript.

If $x \in A_i$ and $y \in B_i$, with $i > 1$, then $\{x, y\}$ separates all of the A's and B's of subscript less than i from those of subscript greater than i. In fact, as depicted in FIGURE 2, an arc can be passed through the holes created by removing x and y, which separates the sets of small subscript from those of large.

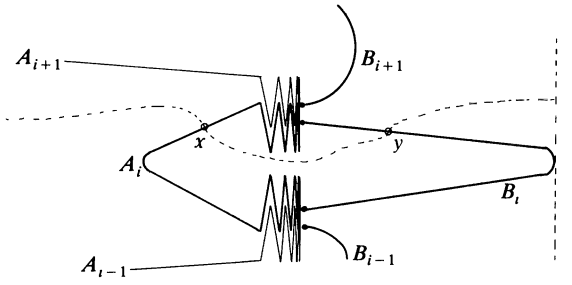


FIG. 2

Consequently, if E meets A_j or B_j, then for each $i > j$, either $A_i \subset E$ or $B_i \subset E$. The same holds true for F.

Therefore, there is a positive integer N such that for $i > N$, one of A_i or B_i is in E and the other is in F.

It is also clear that for each i, the set $A_i \cup A_{i+1}$ separates B_i from B_{i+1} in $X \cup Y$. Therefore, when $i > N$, neither set E nor F can contain both A_i and A_{i+1}, for the other set would then be a connected subset of $(X \cup Y) - (A_i \cup A_{i+1})$ which contains $B_i \cup B_{i+1}$. Of course, this also shows that neither set E nor F can contain both B_i and B_{i+1}.

Since E meets A's of arbitrarily large odd subscript, we can now conclude that there is an integer n such that E contains

$$A_{2n+1} \cup B_{2n+2} \cup A_{2n+3} \cup B_{2n+4} \cup \cdots .$$

This contradicts the fact that F meets B's of arbitrarily large even subscript.

The author has recently shown [2] that for a crossing $[X, Y]$ to be level, one of the sets X or Y must have infinitely many constituents (maximal continuumwise connected subsets). It is also shown that if each of the sets X and Y contains a set which is irreducibly connected between its two corners, then the crossing is not level.

In our example, the sets X and Y each have a countable infinitude of constituents. Also, Y is irreducibly connected between q and q' (there is no smaller connected set containing both of these points), but X does not contain any set which is irreducibly connected between p and p'.

REFERENCES

1. R. Dawson, Paradoxical connections, *American Mathematical Monthly*, 96 (1989) 31–33.
2. K. Whittington, A level crossing of the square requires infinitely many constituents, submitted to *Houston J. Math.*, August 1989.

Lines of Best Fit by Graphics and the Wald Line

R. J. DUFFIN
Department of Mathematics, Carnegie-Mellon University, Pittsburgh, PA 15213

I. Introduction. Here is an old, important and somewhat controversial question: Given a set S of n points in the (x, y) plane and given a straight line then what is the line's disparity? Of course this question is ambiguous on several counts. But as a first approach define "disparity" to be the greatest distance that a point of the set is from the line. Let us denote this distance as D. An example set S is shown in FIGURE 1.

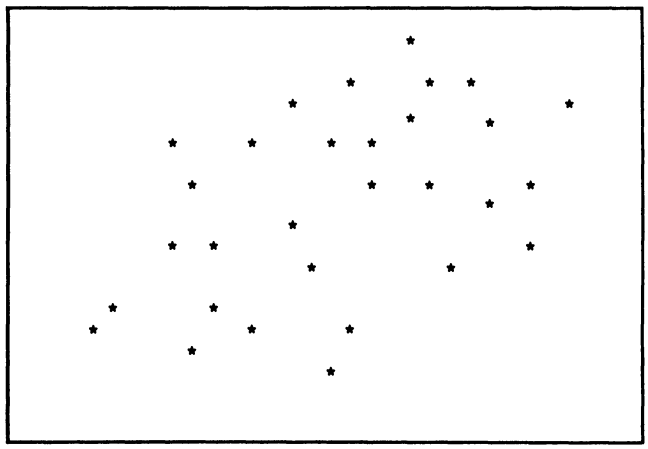


FIG. 1. A set of 31 points to be fitted by a line.

Then another question arises: If the line can vary what is the minimum value of the distance D? Let us denote this minimum as D^*. Let a line of best D fit serve as the midline of an infinite strip of width $2D^*$. Then it is clear that this strip must contain every point of the set S. Next consider any other infinite strip which contains the set S. If its width is W the midline of this strip will have disparity $D \leqslant W/2$. Thus $W \geqslant 2D^*$. The problem of finding D^* is thereby reduced to finding the thinnest strip containing the points S.

II. An algorithm to minimize the distance disparity. A graphical algorithm begins by plotting the n points of S in the (x, y) plane as shown in FIGURE 1. The points of S are contained in a convex polygon whose vertices are all points of S.

Dedicated to a master educator, L. A. Karlovitz (1936–1990).

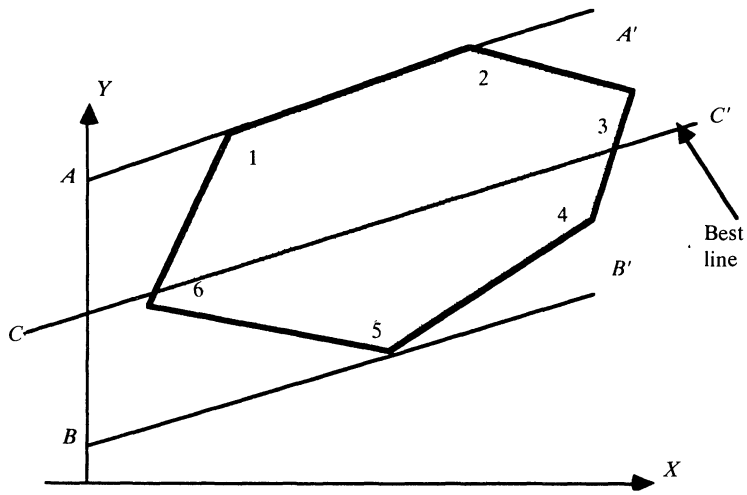


FIG. 2. The fence around the 31 points.

FIGURE 2 shows such a polygonal fence constructed for the set of 31 points shown in FIGURE 1. This polygon has only 6 vertex points.

It is geometrically obvious that no matter where the given line is drawn, there will be a vertex point on the fence which is at a greater distance from the line than any point in the interior of the polygon. For this reason the interior points are discarded. Like empty cans they have served their purpose and are put in the trash barrel. This graphical device greatly simplifies the calculations to be made.

Next a straight line is passed through each edge of the fence polygon. For example, FIGURE 2 shows a line AA' passing through edge $(1, 2)$. Then the vertex of the polygon is selected which is at the greatest distance from the line AA'. In FIGURE 2 this is seen to be the vertex 5. Then a line BB' is passed through vertex 5 and parallel to AA'. The strip formed by the parallel lines AA' and BB' contains all the points of the polygon.

Such a strip is constructed for each edge of the polygon. Each strip has a width. Then a strip with the minimum width W is selected. By examining FIGURE 2, your eye should detect that the edge $(1, 2)$ generates the strip of minimum width W. The other edges are seen to generate strips which are wider.

Then a line CC' is drawn parallel to the lines AA' and BB' but half way between them. By using Lemma 1 to follow it is seen that $D^* = W/2$ and that the midline CC' is an optimum line for best D fit.

LEMMA 1. *Suppose an infinite strip contains a prescribed convex polygon, and that its width is minimal among all such strips with this property. Then each edge of such a strip must contain at least one vertex of the polygon and one of the edges must contain a side of the polygon.*

Proof. If an edge does not contain a vertex the edge can be moved closer to the other edge and the width will be decreased. Thus it can be supposed that each edge contains at least one vertex. Suppose that each edge contains exactly one

vertex of the polygon. Then rotating the strip slightly about the midpoint of these two vertex points results in a strip of the same width which contains a polygon in its interior. This is a contradiction. Thus one edge must contain a side.

III. An algorithm to minimize the vertical disparity. Now consider a second approach to the original question. Instead of taking the disparity of line and point as the distance between them we take it to be the vertical distance between them. Thus the disparity of a line to a set of points is the greatest vertical distance of the points to the line. Let this distance be denoted as V. Fortunately, the value of V is simply related to the value of D. Thus by geometry

$$V = D\sqrt{(1 + m^2)}, \tag{1}$$

where m is the slope of the line. For example consider the V of the line CC' in FIGURE 2. The width of the strip formed from the lines AA' and BB' is W. This strip intercepts the y-axis in a line segment AB having length $|AB| = W\sqrt{(1 + m^2)}$ and consequently

$$V = |AB|/2. \tag{2}$$

If the line is permitted to vary let the minimum V be denoted as V^*. Then (2) gives

$$V^* = \text{Minimum } |AB|/2, \tag{3}$$

where the minimum is taken over all strips formed from the edges of the fence polygon. This is a consequence of Lemma 2 which follows. Formula (3) is readily evaluated graphically. From FIGURE 2, your eye should detect that line CC' is also the best V fit, and so $V^* = |AB|/2$.

LEMMA 2. *Suppose an infinite strip contains a prescribed convex polygon, and that the length $|AB|$ of its y-intercept is minimal among all strips with this property. Then each edge of such a strip must contain at least one vertex of the polygon. Moreover there is one such strip with an edge containing a side of the polygon.*

Proof. Otherwise each edge of a strip of minimum $|AB|$ has exactly one vertex point. Without loss of generality it can be assumed that the lower edge contains the vertex $(0, 0)$ and the upper edge contains the vertex (x', y'). Any new line going through the point (x', y') and having a small difference in angle could serve as an upper edge for a strip containing the whole polygon. If $x' \neq 0$ then it is obvious that the intercept on the y axis can be decreased. This contradiction shows that $x' = 0$. In that case the angle of the new edge does not change the y-intercept. Clearly it can be chosen so that one of the edges contains a side of the polygon.

IV. Relation to linear programs. It is easy to see that the following linear program yields the minimum vertical disparity for a line, $y = mx + b$.

Program 1. Seek $V^* = \min V$ subject to

$$-V \leqslant mx_i + b - y_i \leqslant V, \qquad i = 1, \ldots, n.$$

Thus if S has 31 points there are 62 inequality constraints.

Linus Schrage treats an example problem of fitting a line to five points using linear programming in his business school text book [1]. Our graphic solution to his problem is obvious without need to use the simplex algorithm on a linear program with ten inequality constraints.

Now consider the minimum distance D^*. Relation (1) above, gives

Program 2. Seek $D^* = \min D$ subject to

$$D\sqrt{(1 + m^2)} \leqslant mx_i + b - y_i \leqslant D\sqrt{(1 + m^2)}, \qquad i = 1, \ldots, n.$$

This is neither a linear program nor a quadratic program. So graphics may be the handiest answer to the original question.

V. Is the method of least squares a virtue or a vice? Consider the problem fitting a line to experimental data points: $(x_i, y_i), (x_2, y_2), \ldots, (x_n, y_n)$ in order to determine an underlying physical law, $Y = mX + b$. The trouble is that the data can have experimental errors. A common procedure is to minimize a mean of all vertical disparities. For example, the mean can be the L_p norm of the disparities. For $p = 2$ Gauss found that explicit formulae are readily available. The $p = 2$ method is called *least squares*.

The genius of Gauss notwithstanding, least squares has had difficulty with "outliers". Thus quoting from a book by expert statisticians, [2, page 129], "A wild data point can easily seize control of the fitted-line and cause it to give a totally misleading summary of the relationship between y and x." In his book, Linus Schrage [1] compares the L_1, the L_2 and the L_∞ line fitting methods. (L_∞ gives Program 1.) Schrage argues that the L_1 norm with linear programming is better suited to handle outliers than the traditional least squares method.

VI. The Wald center of gravity method. By far the simplest method to fit a line to experimental data is the Wald formula advocated by the distinguished statistician, Abraham Wald [2], [3]. *His method requires no minimization.* It involves only concepts which can be explained to elementary students and made:

(1) understandable, (2) reasonable and (3) easy to calculate. Here is a scenario of such a lesson:

Students please imagine that there are n points marked on a drawing board, say $(X_1, Y_1), \ldots, (X_n, Y_n)$. It is known that these points all lie on a straight line, however the numerical values of the coordinates are not known. So you ask a skilled technician to measure the coordinates of these n points. He gives you the following results of his measurements: $(x_1, y_1), \ldots, (x_n, y_n)$. Then how do you find the equation of the line?

Of all curves, only the straight line has the property that the center of gravity of any n points on the curve is also on the curve. Let

$$X' = (X_1 + \cdots + X_n)/n \quad \text{and} \quad Y' = (Y_1 + \cdots + Y_n)/n$$

so the point (X', Y') is on the line. Likewise the center of gravity (x', y') of the experimental data is defined by

$$x' = (x_1 + \cdots + x_n)/n \quad \text{and} \quad y' = (x_1 + \cdots + x_n)/n.$$

Your technician uses a measurement procedure with the virtue that a negative error is just as probable as a positive error. Then for large n it is to be expected

that the negative errors will tend to cancel the positive errors so as n becomes large

$$X' - x' \quad \text{and} \quad Y' - y' \text{ both become small.}$$

Thus the point (x', y') is a good *estimator* of the point (X', Y'). You can use it as the "center point" of your approximate line. (*A least squarer would have to agree to this point*!)

However, it takes two points to determine a line so two estimators are needed. Then two, seemingly conflicting, requirements arise. On the one hand all the points should be used because that is how the errors can cancel. On the other hand the two estimator points should be far apart because the true line and the approximate line will stay close between the estimator points.

To meet these requirements start over. Suppose that actually there were $2n$ points on the line. Then split the (x, y) measurements into two political parties: "leftists" and "rightists." This reordering makes

$$x_1, x_2, \ldots, x_n \quad \text{all less than} \quad x_{n+1}, \ldots, x_{2n}.$$

Then let

$$X'' = (X_{n+1} + \cdots + X_{2n})/n \quad \text{and} \quad Y'' = (Y_{n+1} + \cdots + Y_{2n})/n.$$

Thus (X'', Y'') is a point on the line and its *estimator* (x'', y'') is so defined.

The point (x', y') is the center of gravity of the Leftists and (x'', y'') is the center of gravity of the Rightists. The line which goes through these two points is the **Wald Line**. Then, of course,

$$(\mathbf{y - y'}) \,/\, (\mathbf{x - x'}) = (\mathbf{y'' - y'}) \,/\, (\mathbf{x'' - x'})$$

is the **Wald Formula**. That, in essence, is all there is to it.

As homework problems for next time: Show that

(1) The center of gravity of all the data points is on the Wald line.
(2) The disparity between the Wald line and the true line in the interval (x', x'') does not exceed the disparity at the estimator points.
(3) If the x values of the data points are distributed rather uniformly in the whole interval then the disparity at any point does not exceed twice the disparity at the estimator points.

The data enter linearly in the Wald formula and not quadratically as in the traditional least squares formula. Thus the Wald line is much less disturbed by wild data points than the least squares line. However least squarers are probably not familiar with the Wald Line. When they are told about it their first question is, "But what does it minimize?"

REFERENCES

1. Linus Schrage, Linear, Integer and Quadratic Programming with *LINDO*, chapter 15, The Scientific Press, Palo Alto, CA (1988).
2. David C. Hoaglin, Frederick Mosteller, and John W. Tukey, Understanding Robust and Exploratory Data Analysis, John Wiley, New York, 1983.
3. Abraham Wald, The fitting of straight lines if both variables are subject to error, *Annals of Math Statistics*, 11 (1940) 284–300.

On the Mean Value Theorem, Inequality, and Inclusion

M. Furi
Dipartimento di Matematica Applicata, Università di Firenze, Via S. Marta 3, 50139 Firenze, Italy

M. Martelli
Mathematics Department, California State University, Fullerton, CA 92634

The Mean Value or Lagrange's* Theorem and its consequences are one of the fundamental topics of a calculus course. The theorem is usually presented in the following form.

Theorem 1. *Let $f:[a,b]\to\mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c\in(a,b)$ such that*

$$f(b)-f(a)=f'(c)(b-a). \tag{1}$$

The result is frequently derived from the following theorem, which is credited to Rolle†.

Theorem 2. *Let $f:[a,b]\to\mathbb{R}$ be continuous on $[a,b]$, differentiable on (a,b) and such that $f(a)=f(b)$. Then there exists $c\in(a,b)$ such that*

$$f'(c)=0. \tag{2}$$

The simplest way of obtaining (1) from (2) is to start from the desired result

$$f'(c)-\frac{f(b)-f(a)}{b-a}=0.$$

The left side of this equality is the derivative at c of the function

$$g(x)=f(x)-\frac{f(b)-f(a)}{b-a}x.$$

Since it can be easily seen that $g(a)=g(b)$, the second part of the equality follows from Rolle's Theorem. Moreover (2) is easily derived from (1) when $f(a)=f(b)$. Hence the two results are equivalent.

At the level of advanced calculus, students find that the Mean Value Theorem is not valid in the simple form expressed by (1), for vector-valued maps. One of the simplest examples to show this is provided by $f(t)=(\cos t,\sin t)$. Since $f(0)=$

*Joseph Louis Lagrange attempted a reorganization of calculus by representing functions with their Taylor's series expansions. The Mean Value Theorem is frequently attributed to him, particularly in European textbooks. However the first statement of the theorem appears in a paper [2] of the renowned physicists André-Marie Ampère (1775–1836). In the paper Ampère uses the ideas of Lagrange to justify the result. The derivative is assumed to be continuous. About fifteen years later Augustin-Louis Cauchy (1789–1857) proved the theorem with the same hypothesis (see [3, p. 113]).

It is worth mentioning that in the same paper Ampère claims to have proved a result that today we know to be false, namely that a continuous function is differentiable except possibly at finitely many exceptional points. The stature of the man who felt so strongly about this false result to the point of proving it, tells us that scientists have fought an uphill battle to bring calculus to today's form.

†Michel Rolle (1652–1719) was a member of the French Academy and engaged in a heated debate on the foundations of "infinitesimal methods," saying that these methods led to paralogisms. He was rebutted by John Bernoulli who maintained that Rolle did not understand the calculus.

$f(2\pi)$ and

$$\|f'(t)\| = \sqrt{\sin^2 t + \cos^2 t} = 1$$

it is obvious that (1) is violated.

At this point many authors prove that Lagrange's Theorem can be extended to scalar functions of vector variables (see, e.g., [5, p. 174]). A few authors go ahead and present a result called the Mean Value Inequality (see, e.g., [4, p. 379]). To state the result as simply as possible we shall use the symbol $[\mathbf{x}, \mathbf{y}]$ $(]\mathbf{x}, \mathbf{y}[)$ to denote the elements of the closed (open) line segments joining $\mathbf{x}$ with $\mathbf{y}$. Moreover we shall use the Euclidean norm in $\mathbb{R}^n$.

THEOREM 3. *Let $U \subset \mathbb{R}^n$ be open and $f: U \to \mathbb{R}^m$ be differentiable on U with continuous derivative. Let $\mathbf{x}, \mathbf{y} \in U$ be such that $[\mathbf{x}, \mathbf{y}] \subset U$. Then*

$$\|f(\mathbf{y}) - f(\mathbf{x})\| \leqslant \max\{\|f'(\mathbf{c})\| : \mathbf{c} \in [\mathbf{x}, \mathbf{y}]\}\|\mathbf{y} - \mathbf{x}\|, \tag{3}$$

where $\|f'(\mathbf{c})\| = \max\{\|f'(\mathbf{c})\mathbf{x}\| : \|\mathbf{x}\| = 1\}$.

A comparison between the Mean Value Inequality and Lagrange's Theorem shows that the Inequality, as stated above, is based on assumptions a bit stricter than the ones used in the Theorem, although the conclusion is not as sharp. Theorem 3 is frequently proved by using, among other things, the Fundamental Theorem of Integral Calculus (see [1, p. 7; 4, p. 379]).*

Our goal in this paper is to show that the Mean Value Inequality can be established in a more elementary manner, and under assumptions that are weaker than the ones used in the Mean Value Theorem. Moreover, we shall obtain a result, called the Mean Value Inclusion, which can be regarded as the form of Lagrange's Theorem [the Mean Value Theorem] in dimension higher than 1.

In our final remarks, we shall point out that the differentiability of f can be replaced by a less restrictive condition, called Gâteaux differentiability. The two notions coincide in $\mathbb{R}$, but differ in higher dimension.

Consequences of the Mean Value Theorem. The following two corollaries of the Mean Value Theorem are usually included in textbooks. We mention them here because, as we shall see, the Mean Value Inequality provides their extension to vector functions.

COROLLARY 1. *Let $f: [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a,b). Assume that $f'(x) = 0$ for all $x \in (a, b)$. Then f is constant.*

COROLLARY 2. *Let $I \subset \mathbb{R}$ be an interval and $f: I \to \mathbb{R}$ be continuous on I and differentiable on the interior of I. Assume that f' is bounded. Then f is uniformly continuous on I.*

*W. Rudin, in the first and second edition of [8] gives a different, more elaborate proof of a special case of (3). A shorter proof was presented by V. P. Havin, when he translated the book into Russian and it was added to the third edition of Rudin's book. Our proof is along the lines of the one of Havin.

To prove Corollary 1 we choose any point $x \in (a, b]$ and apply the Mean Value Theorem to f on $[a, x]$. We obtain

$$f(x) - f(a) = f'(c)(x - a)$$

for some $c \in (a, x)$. Since $f'(c) = 0$, it follows that $f(x) = f(a)$.

To establish Corollary 2 we select any pair of points $x, y \in I$, $x < y$. From the Mean Value Theorem applied to f on $[x, y]$ and the boundedness of f', we obtain

$$|f(y) - f(x)| \leqslant M|y - x|$$

for some constant $M \geqslant 0$. This inequality implies the uniform continuity of f.

We now present two more consequences of the Mean Value Theorem. They are not included in most textbooks. The reason for having them here is to simplify the proof of the Mean Value Inequality and of the Mean Value Inclusion. One could argue that the two results have also some interest on their own.

COROLLARY 3. *Let $f:[a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a,b) except possibly at finitely many points. Then there exists $c \in (a, b)$ such that*

$$|f(b) - f(a)| \leqslant |f'(c)|(b - a). \tag{4}$$

Proof. Assume that there is only one point $d \in (a, b)$ where f is not differentiable. By applying the Mean Value Theorem to f on $[a, d]$ and $[d, b]$ respectively we obtain

$$f(d) - f(a) = f'(c_1)(d - a), \qquad c_1 \in (a, d),$$

$$f(b) - f(d) = f'(c_2)(b - d), \qquad c_2 \in (d, b).$$

From these equalities we derive

$$|f(b) - f(a)| \leqslant |f'(c)|(b - a)$$

where

$$|f'(c)| = \max\{|f'(c_1)|, |f'(c_2)|\}.$$

This proof can obviously be extended to the case when f is not differentiable at more than one point. □

COROLLARY 4. *Let $f:[a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a,b) except possibly at a finite number, n, of points. Then there are $n + 1$ points $a < c_1 < c_2 < \cdots < c_{n+1} < b$ and $n + 1$ positive numbers $\alpha_1, \ldots, \alpha_{n+1}$ such that*

$$\sum_{i=1}^{n+1} \alpha_i = 1$$

and

$$f(b) - f(a) = \sum_{i=1}^{n+1} \alpha_i f'(c_i)(b - a). \tag{5}$$

Proof. Assume that there is only one point $d \in (a, b)$ where f is not differentiable. By applying the Mean Value Theorem to f on $[a, d]$ and $[d, b]$ respectively, we obtain

$$f(d) - f(a) = f'(c_1)(d - a)$$

$$f(b) - f(d) = f'(c_2)(b - d)$$

for some $c_1 \in (a, d)$, $c_2 \in (d, b)$. Let α_1, α_2 be such that

$$\alpha_1(b - a) = d - a$$
$$\alpha_2(b - a) = b - d.$$

Then $\alpha_1 + \alpha_2 = 1$, $\alpha_1 > 0$, $\alpha_2 > 0$ and

$$f(b) - f(a) = (\alpha_1 f'(c_1) + \alpha_2 f'(c_2))(b - a).$$

This proof can obviously be extended to the case when f is not differentiable at $n > 1$ points.

The Mean Value Inequality and Inclusion. We now present the Mean Value Inequality in a form that is more general than the one given in Theorem 3. We also include a particular form of the Mean Value Inclusion (Theorem 5). For an extensive analysis of this last result the reader should consult [**6**]. To simplify both statements we introduce the following:

Definition. A subset V of $\mathbb{R}^n$ is *linearly finite* if for every pair of points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the set

$$V \cap [\mathbf{x}, \mathbf{y}]$$

is finite.

THEOREM 4 (Mean Value Inequality). *Let D be a disk in $\mathbb{R}^n$ and $f: D \to \mathbb{R}^m$ be continuous on D. Assume that f is differentiable on the interior, B, of D, except possibly on a linearly finite subset of B. Let $\mathbf{x}, \mathbf{y} \in D$. Then there exists $\mathbf{c} \in]\mathbf{x}, \mathbf{y}[$ such that*

$$\|f(\mathbf{y}) - f(\mathbf{x})\| \leqslant \|f'(\mathbf{c})\| \, \|\mathbf{y} - \mathbf{x}\|. \tag{6}$$

Recall that the convex closure of A, $\overline{\text{co}}A$, where A is any subset of $\mathbb{R}^n$, is the closure of the set of all convex linear combinations of elements of A.

THEOREM 5. (Mean Value Inclusion). *Let D be a disk in $\mathbb{R}^n$ and $f: D \to \mathbb{R}^m$ be continuous on D. Assume that f is differentiable on the interior, B, of D, except possibly on a linearly finite subset V of B. Let $\mathbf{x}, \mathbf{y} \in D$. Then*

$$f(\mathbf{y}) - f(\mathbf{x}) \in \overline{\text{co}}\, A, \tag{7}$$

where $A = \{f'(\mathbf{c})(\mathbf{y} - \mathbf{x}) : \mathbf{c} \in V^c \cap]\mathbf{x}, \mathbf{y}[\}$.

For completeness we shall state first, without proofs, all results needed to establish Theorems 4 and 5.

Cauchy–Schwarz Inequality (see [**4**, p. 20]). Let $\mathbf{u}, \mathbf{v}$ be two vectors of $\mathbb{R}^n$ and denote by $\langle \mathbf{u}, \mathbf{v} \rangle$ their inner product. Then the well-known Cauchy–Schwarz inequality states that

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leqslant \|\mathbf{u}\| \, \|\mathbf{v}\|. \tag{8}$$

Differentiability. Let $U \subset \mathbb{R}^n$ be an open set, $f: U \to \mathbb{R}^m$ and $\mathbf{c} \in U$. We say that f is differentiable at $\mathbf{c}$ if there exists a linear operator $L: \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(\mathbf{c} + \mathbf{h}) = f(\mathbf{c}) + L\mathbf{h} + o(\mathbf{h}), \tag{9}$$

where

$$\lim_{\|\mathbf{h}\| \to 0} \frac{o(\mathbf{h})}{\|\mathbf{h}\|} = 0.$$

The linear operator L is usually denoted by $f'(\mathbf{c})$ and, if we define

$$\|L\| = \max\{\|L\mathbf{x}\| : \|\mathbf{x}\| = 1\}$$

then

$$\|L\mathbf{x}\| \leqslant \|L\| \, \|\mathbf{x}\| \tag{10}$$

for all $\mathbf{x} \in \mathbb{R}^n$ (see [**8**, p. 208]).

Chain rule. Let $f: U \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g:[a, b] \to U$ be differentiable on the open set U and on the open interval (a, b) respectively. Then the map $h: [a, b] \to \mathbb{R}^m$ defined by $h(t) = f(g(t))$ is differentiable on (a, b) and its derivative at $t \in (a, b)$ can be identified with the vector obtained by applying the linear operator $f'(g(t))$ to the vector $g'(t)$, i.e.,

$$h'(t) = f'(g(t))(g'(t)). \tag{11}$$

Moreover, for every fixed vector $\mathbf{z} \in \mathbb{R}^m$, the derivative of the map $k:[a, b] \to \mathbb{R}$, defined by $k(t) = \langle \mathbf{z}, h(t) \rangle$ is

$$k'(t) = \langle \mathbf{z}, h'(t) \rangle. \tag{12}$$

Inequalities (8), (10) and equalities (11), (12) are used, in conjunction with Theorem 1 and with the Fundamental Theorem of Integral Calculus, to establish Theorem 3 (see [**1**, p. 7]; [**4**, p. 379]). Our proof of Theorem 4, which is more general than Theorem 3 uses only (8), (10), (11), (12), and Corollary 1.

A separation theorem. We complete our list with a separation theorem needed in the proof of the Mean Value Inclusion. We shall use it in its simplest form.

THEOREM 6. (Separation Theorem). *Let C be a convex closed subset of $\mathbb{R}^m$ and let $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} \notin C$, be a fixed vector. Then there exists a vector $\mathbf{w} \in \mathbb{R}^m$, and a real number a such that*

$$\langle \mathbf{w}, \mathbf{x} \rangle < a \leqslant \langle \mathbf{w}, \mathbf{y} \rangle \quad \textit{for all } \mathbf{y} \in C. \tag{13}$$

From a geometrical point of view, the above theorem says that a closed, convex subset C of $\mathbb{R}^m$ and a vector $\mathbf{x} \notin C$, can be separated by a hyperplane, i.e., there exists a hyperplane H such that C cannot be reached from $\mathbf{x}$ without crossing H (see [**7**, p. 99]).

Proof of the Mean Value Inequality and of the Mean Value Inclusion. We are now ready to prove Theorems 4 and 5.

Proof of Theorem 4. Let $\mathbf{z}$ be a fixed vector of $\mathbb{R}^m$, which will be specified later. Define

$$k:[0, 1] \to \mathbb{R} \quad \text{by } k(t) = \langle \mathbf{z}, f((1 - t)\mathbf{x} + t\mathbf{y}) \rangle.$$

It is easily seen that k satisfies all assumptions of Corollary 3. Hence, by (4),

$$|k(1) - k(0)| \leqslant |k'(t_0)|.$$

for some $t_0 \in (0, 1)$. Since

$$k(1) - k(0) = \langle \mathbf{z}, f(\mathbf{y}) - f(\mathbf{x}) \rangle \quad \text{and} \quad k'(t_0) = \langle \mathbf{z}, f'(\mathbf{c})(\mathbf{y} - \mathbf{x}) \rangle$$

with $\mathbf{c} = (1 - t_0)\mathbf{x} + t_0\mathbf{y}$, we obtain, using inequalities (8) and (10)

$$|\langle \mathbf{z}, f(\mathbf{y}) - f(\mathbf{x}) \rangle| \leqslant \|\mathbf{z}\| \, \|f'(\mathbf{c})\| \, \|\mathbf{y} - \mathbf{x}\|. \tag{14}$$

We can assume $f(\mathbf{y}) \neq f(\mathbf{x})$, for, otherwise inequality (6) is obvious. By selecting

$$\mathbf{z} = \frac{f(\mathbf{y}) - f(\mathbf{x})}{\|f(\mathbf{y}) - f(\mathbf{x})\|}$$

and substituting it into (14), we obtain

$$\|f(\mathbf{y}) - f(\mathbf{x})\| \leqslant \|f'(\mathbf{c})\| \, \|\mathbf{y} - \mathbf{x}\|. \qquad \square$$

The Mean Value Inequality allows us to extend Corollaries 1 and 2 to vector-valued functions. The changes needed in the statements and in the proofs are so minor that we feel unnecessary to explicitly include them here.

We now prove Theorem 5 following the pattern of the proof of Theorem 4.

Proof of Theorem 5. Assume that

$$f(\mathbf{y}) - f(\mathbf{x}) \notin \overline{\text{co}}A. \tag{15}$$

Then, by Theorem 6, there exists a vector $\mathbf{w} \in \mathbb{R}^m$ and a real number a such that

$$\langle \mathbf{w}, f(\mathbf{y}) - f(\mathbf{x}) \rangle < a \leqslant \langle \mathbf{w}, \mathbf{z} \rangle \tag{16}$$

for all $\mathbf{z} \in \overline{\text{co}}A$.

Assume that there are n points of the open line segment joining $\mathbf{x}$ with $\mathbf{y}$, where f is not differentiable. Then Corollary 4 can be applied to $k: [0, 1] \to \mathbb{R}$ defined by

$$k(t) = \langle \mathbf{w}, f((1 - t)\mathbf{x} + t\mathbf{y}) \rangle.$$

We obtain, by (5),

$$k(1) - k(0) = \sum_{i=1}^{n+1} \alpha_i k'(t_i).$$

Hence

$$\langle \mathbf{w}, f(\mathbf{y}) - f(\mathbf{x}) \rangle = \left\langle \mathbf{w}, \sum_{i=1}^{n+1} \alpha_i f'(\mathbf{c}_i)(\mathbf{y} - \mathbf{x}) \right\rangle$$

where $\mathbf{c}_i = (1 - t_i)\mathbf{x} + t_i\mathbf{y}$. This last equality contradicts (15). $\square$

We have stated and proved Theorems 4 and 5 for vector-valued functions which are differentiable in the sense of (9), a property usually called Fréchet differentiability. The two results can be extended with no changes to functions $f: D \subset \mathbb{R}^n \to \mathbb{R}^m$ which satisfy a weaker definition of differentiability, called Gâteaux differentiability.

We shall say that f is Gâteaux differentiable (see [**1**, p. 5]; [**9**, p. 11]) at an interior point $\mathbf{x} \in D$ if there is a linear operator $L: \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{h}) - f(\mathbf{x})}{t} = L\mathbf{h} \tag{17}$$

for all $\mathbf{h} \in \mathbb{R}^n$.

While the composition of two functions that are Gâteaux differentiable need not be Gâteaux differentiable, it can be seen that (11) still holds when the Fréchet differentiability is replaced by the Gâteaux differentiability.

In fact let $\mathbf{u} = \alpha(t) = (1 - t)\mathbf{x} + t\mathbf{y}$, $t \in (0, 1)$ and let $f'_G(\mathbf{u})$ be the Gâteaux derivative of f at $\mathbf{u}$. Let s be such that $t + s < 1$. Then $f(\alpha(t + s)) = f(\alpha(t) + s(\mathbf{y} - \mathbf{x}))$ and

$$\lim_{s \to 0} \frac{f(\alpha(t + s)) - f(\alpha(t)) - sf'_G(\mathbf{u})(\mathbf{y} - \mathbf{x})}{s} = 0, \tag{18}$$

which shows that $f \circ \alpha$ is Gâteaux differentiable at t and its Gâteaux derivative is $f'_G(\mathbf{u})(\mathbf{y} - \mathbf{x})$.

The Gâteaux differentiability is strictly more general than the Fréchet differentiability as the following example shows.

Example 1. Let

$$f(x, y) = \begin{cases} \dfrac{y^3}{\sqrt{x^2 + y^4}} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

Then f is continuous at $(0, 0)$ and its Gâteaux derivative at $(0, 0)$ is the zero linear operator. However f is not Fréchet differentiable at $(0, 0)$ as can be checked by approaching the origin along the parabola $x = y^2$.

Theorems 4 and 5 can be extended to infinite-dimensional Hilbert spaces and to Banach spaces. In a Hilbert space H the proofs given before for the finite dimensional setting do not need any changes. In a Banach space E, however, there is no inner product compatible with the norm, and one has to use the Hahn–Banach theorem to prove the existence of a functional $\phi: E \to \mathbb{R}$ such that $\|\phi\| = 1$ and $\phi(f(\mathbf{x}) - f(\mathbf{y})) = \|f(\mathbf{x}) - f(\mathbf{y})\|$.

Acknowledgments. We would like to thank the referee for many helpful suggestions, which have improved the presentation of this paper.

REFERENCES

1. A. Ambrosetti, G. Prodi, Analisi Nonlineare, Quaderni della Scuola Normale Superiore di Pisa, I Quaderno, Pisa, 1973.
2. A. M. Ampère, Rechèrche sur quelques points de la théorie des functions, *J. Ecole Polyt., Cah.* 13, 6 (1806), 148–181.
3. I. Grattan-Guiness, From the Calculus to Set Theory, 1630–1910, Duckworth, London, 1980.
4. S. Lang, Undergraduate Analysis, Springer-Verlag, New York, 1983.
5. J. Marsden, Elementary Classical Analysis, W. H. Freeman, San Francisco, 1974.
6. R. McLeod, Mean Value Theorems for vector-valued functions, *Proc. Edin. Math. Soc.* 14 (1965), 197–209.
7. R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1980.
8. W. Rudin, Principles of Mathematical Analysis, 3rd edition, McGraw-Hill, New York, 1976.
9. J. T. Schwartz, Nonlinear Functional Analysis, Gordon and Breach Science Publishers, New York, 1969.

The "big oh" notation means that there is a constant C such that

$$|D_h f(x) - f'(x)|/h \leqslant C.$$

In fact, if we set

$$\|D_h f - f'\| = \sup_{x \in I} |D_h f(x) - f'(x)|,$$

then

$$\|D_h f - f'\| = O(h).$$

Suppose now that only an approximate version of f, f_η, having the form

$$f_\eta(t) = f(t) + \eta(t)$$

is known, where η is a bounded, but not necessarily differentiable, function representing the error in the approximation. The approximation to the derivative of f is then given by

$$D_h f_\eta(x) = D_h f(x) + D_h \eta(x). \tag{2}$$

Since η is not necessarily differentiable, $\lim_{h \to 0} D_h \eta(x)$ does not necessarily exist. Even in the case when η is differentiable, it can happen, as we saw earlier, that this limit can be very large even when $\|\eta\|$ is very small.

How bad can things get? Suppose that $\|\eta\| \leqslant \varepsilon$, where $\varepsilon > 0$ is a positive number giving a bound for the error. Then from (1) and (2) we have

$$\begin{aligned} \|D_h f_\eta - f'\| &\leqslant O(h) + \|D_h \eta\| \\ &\leqslant O(h) + \frac{2\varepsilon}{h}. \end{aligned} \tag{3}$$

Note that for fixed $\varepsilon > 0$, the error bound blows up as $h \to 0$. Nevertheless, the beast can, in a sense, be tamed or at least regulated. We simply cannot let $h \to 0$ in an arbitrary manner, but must choose a strategy of relating h to ε in an appropriate way. In fact, we see that any choice of $h(\varepsilon) > 0$ which satisfies $h(\varepsilon) \to 0$ and $\varepsilon/h(\varepsilon) \to 0$ as $\varepsilon \to 0$ results in

$$\|D_{h(\varepsilon)} f_\eta - f'\| \to 0 \quad \text{as } \varepsilon \to 0.$$

The forward difference approximation method with such a choice of grid size $h = h(\varepsilon)$ is then said to provide a regular algorithm in the sense of Tikhonov [2] for the problem of differentiation.

From (3) we see that a choice of grid size of the form $h(\varepsilon) = C\sqrt{\varepsilon}$ (obtained by "balancing" the terms $O(h)$ and $2\varepsilon/h$ in (3)) leads to a $\sqrt{\varepsilon}$ order of approximation:

$$\|D_{h(\varepsilon)} f_\eta - f'\| = O(\sqrt{\varepsilon})$$

for all perturbations η satisfying $\|\eta\| \leqslant \varepsilon$.

Is it possible to do better? Obviously if f is a linear function, then $D_h f = f'$ and hence a stronger $o(\sqrt{\varepsilon})$ order of approximation is attainable for any choice of grid size of the form $h(\varepsilon) = \varepsilon^a, 0 < a < 1/2$ (recall that $m(\varepsilon) = o(n(\varepsilon))$ means $|m(\varepsilon)/n(\varepsilon)| \to 0$ as $\varepsilon \to 0$). Our aim is to show that only in this trivial case is a $o(\sqrt{\varepsilon})$ order of approximation attainable for all perturbations. The theorem that we prove

below is akin to a result of Bajšanski and Bojanić [1] on the impossibility of improving a "O" Bernstein polynomial approximation to a "o" approximation, except in the linear case.

THEOREM. *Suppose that for some* $h = h(\varepsilon) \to 0$, $\|D_{h(\varepsilon)}f_\eta - f'\| = o(\sqrt{\varepsilon})$ *for every perturbation* η *with* $\|\eta\| \leqslant \varepsilon$. *Then* f *is a linear function.*

Proof. We show that $f''(x) = 0$ for every $x \in I$. By Taylor's theorem,

$$f(x + h) = f(x) + f'(x)h + \int_x^{x+h} f''(t)(x + h - t)\, dt.$$

Therefore,

$$D_h f(x) - f'(x) = R(h),$$

where

$$R(h) = \frac{1}{h}\int_x^{x+h} f''(t)(x + h - t)\, dt.$$

Suppose that for some $x \in I$, $f''(x) = 2\delta > 0$. Then $f''(t) \geqslant \delta > 0$ for $t \in [x, x + h]$, if h is sufficiently small. It follows that $R(h) \geqslant \delta h/2$. Define a perturbation $\eta(t)$ for $t \in I$ by

$$\eta(t) = \begin{cases} \varepsilon, & t > x \\ 0, & t \leqslant x. \end{cases}$$

Then

$$D_h f_\eta(x) - f'(x) = R(h) + \varepsilon/h \geqslant \delta h/2 + \varepsilon/h.$$

It follows, since $\|D_{h(\varepsilon)}f_\eta - f'\| = o(\sqrt{\varepsilon})$, that

$$\frac{\delta}{2}\frac{h(\varepsilon)}{\sqrt{\varepsilon}} + \frac{\sqrt{\varepsilon}}{h(\varepsilon)} \to 0 \quad \text{as } \varepsilon \to 0,$$

which is clearly impossible. A similar argument shows that it is not possible for $f''(x)$ to be negative. Therefore $f''(x) = 0$ for all $x \in I$, and the proof is complete.

Similar results can be proved for other approximations to the derivative. The reader is invited to show, for example, that the centered difference approximation $(D_h f(x) + D_{-h}f(x))/2$ and the extrapolated forward difference approximation $2D_h f(x) - D_{2h}f(x)$ for bounded functions with three continuous bounded derivatives on an open interval both lead, for appropriate choice of grid size, to a $O(\varepsilon^{2/3})$ approximation in the presence of perturbations of size ε. Moreover, this order of approximation cannot be improved to $o(\varepsilon^{2/3})$, unless f is a quadratic polynomial.

Of course it is not necessary to limit oneself to the first derivative. Finite difference approximations to higher order derivatives in the presence of errors can be analyzed in a similar way.

Finally, we give a numerical illustration of the choice of grid size and resulting order of convergence for the forward difference approximation of the derivative. The numerical values given below are the results of a simple program in which the perturbation $\eta(x) \in [-\varepsilon, \varepsilon]$ is produced by a (uniform) random number generator and the approximate derivative $D_h f_\eta$ is compared with the true derivative at grid

points in the interval [0, 1]. The grid size is taken as $h = h(\varepsilon) = \sqrt{\varepsilon}$ and therefore the error satisfies

$$\|D_h f_\eta - f'\| = O(\sqrt{\varepsilon}).$$

This behavior can be numerically validated by assuming that the error has the form $E \sim C\varepsilon^p$. Taking ratios, the order of convergence p can then be estimated in terms of two calculated errors E_1 and E_2, corresponding to two perturbation levels ε_1 and ε_2, respectively, by

$$\ln(E_1/E_2)/\ln(\varepsilon_1/\varepsilon_2).$$

For nonlinear functions, we should expect an order $p = 1/2$ at best, according to the theorem. For example, if $f(x) = (2x - 1)e^x$, then the errors calculated in the program and displayed in the output below can be used to estimate p (to two decimals) as shown in the table:

ε	.1		.01		.001		.0001		.00001
E	1.330752		.7109080		.2339649		.07577038		.02447224
p		.27		.48		.50		.49	

REFERENCES

1. B. Bajšanski and R. Bojanić, A note on approximation by Bernstein polynomials, *Bull. Amer. Math. Soc.*, 70 (1964) 675–677.
2. A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-posed Problems, Wiley, New York, 1977 (translated from the Russian).

More on Cross Products

PAUL BINDING*
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

A recent article [1] in the MONTHLY contains several proofs of the equivalent identities

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}) \tag{1}$$

and

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}. \tag{2}$$

Since one motivation for this work is the difficulty of completing a particular geometrical argument for (2), it seems worthwhile to give a simple, direct geometrical proof.

We adopt the geometrical definition that $\mathbf{a} \times \mathbf{b} \perp$ (i.e., is orthogonal to) $\mathbf{a}$ and $\mathbf{b}$, according to the right-hand rule, with

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \, \|\mathbf{b}\| \sin \theta, \tag{3}$$

θ being the angle between $\mathbf{a}$ and $\mathbf{b}$. It follows that

$$\mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b} \tag{4}$$

*Research supported by NSERC of Canada.

points in the interval [0, 1]. The grid size is taken as $h = h(\varepsilon) = \sqrt{\varepsilon}$ and therefore the error satisfies

$$\|D_h f_\eta - f'\| = O(\sqrt{\varepsilon}).$$

This behavior can be numerically validated by assuming that the error has the form $E \sim C\varepsilon^p$. Taking ratios, the order of convergence p can then be estimated in terms of two calculated errors E_1 and E_2, corresponding to two perturbation levels ε_1 and ε_2, respectively, by

$$\ln(E_1/E_2)/\ln(\varepsilon_1/\varepsilon_2).$$

For nonlinear functions, we should expect an order $p = 1/2$ at best, according to the theorem. For example, if $f(x) = (2x - 1)e^x$, then the errors calculated in the program and displayed in the output below can be used to estimate p (to two decimals) as shown in the table:

ε	.1		.01		.001		.0001		.00001
E	1.330752		.7109080		.2339649		.07577038		.02447224
p		.27		.48		.50		.49	

REFERENCES

1. B. Bajšanski and R. Bojanić, A note on approximation by Bernstein polynomials, *Bull. Amer. Math. Soc.*, 70 (1964) 675–677.
2. A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-posed Problems, Wiley, New York, 1977 (translated from the Russian).

More on Cross Products

PAUL BINDING*
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

A recent article [1] in the MONTHLY contains several proofs of the equivalent identities

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}) \tag{1}$$

and

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}. \tag{2}$$

Since one motivation for this work is the difficulty of completing a particular geometrical argument for (2), it seems worthwhile to give a simple, direct geometrical proof.

We adopt the geometrical definition that $\mathbf{a} \times \mathbf{b} \perp$ (i.e., is orthogonal to) $\mathbf{a}$ and $\mathbf{b}$, according to the right-hand rule, with

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \theta, \tag{3}$$

θ being the angle between $\mathbf{a}$ and $\mathbf{b}$. It follows that

$$\mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b} \tag{4}$$

*Research supported by NSERC of Canada.

(Miquel's Theorem) that the circles $AB'C'$, $A'BC'$, $A'B'C$ intersect in a point Y. Prove that $X = Y$ if and only if X is the orthocenter of $\triangle ABC$.

E 3467. *Proposed by Todd Feil, Denison University, Granville, OH, and Gary Kennedy, Oberlin College, Oberlin, OH.*

A permutation π on the set $\{1, 2, \ldots, n\}$ is said to have j as *a strong fixed point* if $\pi(k) < j$ for $k < j$ and $\pi(k) > j$ for $k > j$. Let $h(n)$ be the number of permutations on $\{1, 2, \ldots, n\}$ having at least one strong fixed point. Prove that

$$2(n-1)! - (n-2)! \leqslant h(n) \leqslant 2(n-1)!$$

for $n > 1$.

E 3468. *Proposed by Curtis Cooper, Central Missouri State University, Warrensburg, MO, Robert E. Kennedy, Central Missouri State University, and Stanley Rabinowitz, Westford, MA.*

Suppose m and n are positive integers such that all prime factors of n are larger than m.

(a) Prove that

$$\sum_{k=1}^{n}{}^{*} \sin^{2m}\left(\frac{k\pi}{n}\right) = \frac{\phi(n) - \mu(n)}{4^m}\binom{2m}{m},$$

where * denotes summation over integers relatively prime to n. (Here ϕ and μ denote the arithmetic functions of Euler and Möbius respectively.)

(b) Find a similar formula for

$$\sum_{k=1}^{n}{}^{*} \cos^{2m}(k\pi/n).$$

SOLUTIONS OF ELEMENTARY PROBLEMS

A Bounded Binomial Experiment

E 3332 [1989, 524]. *Proposed by John Mamer, University of California at Los Angeles, and Kenneth Schilling, University of Michigan-Flint.*

Let X be a binomial random variable with parameters n and p.

(a) Prove that

$$E(\min\{X, t\}) \geqslant t\{1 - (1 - p/t)^n\}$$

for $t = 1, 2, \ldots, n$.

(b)* Is there an upper bound for $E(\min\{X, t\})$ better than $\min(np, t)$?

Solution I of (a) independently by Wolfgang J. Buhler, Johannes Gutenberg University, Mainz, Germany, Glenn A. Stoops, Dayton, OH, and the proposers. An archer, whose probability of hitting a chosen target is p, performs the following experiment. For each of n arrows he selects one of t targets at random and fires the arrow at it. For any fixed target, his probability of not hitting it with a

particular arrow is $1 - p/t$ and so his probability of not hitting it at all is $(1 - p/t)^n$. Hence the expected number of targets hit is $t\{1 - (1 - p/t)^n\}$.

On the other hand, if targets already hit are excluded from consideration before each arrow is fired, the expected number of targets hit can only increase. If this modus operandi is followed, the probability of a new hit with any arrow is p unless all targets have already been hit. Thus the expected number of targets hit under this revised procedure is $E(\min\{X, t\})$. Comparing the two expectations, we obtain the inequality given in (a).

Solution II of (a) by K. F. Andersen, University of Alberta, Edmonton, Canada. By definition

$$E(\min\{X,t\}) = \sum_{j=0}^{t-1} \binom{n}{j} p^j (1-p)^{n-j} j + \sum_{j=t}^{n} \binom{n}{j} p^j (1-p)^{n-j} t$$

$$= t\left\{1 - \sum_{j=0}^{t-1} \binom{n}{j} p^j (1-p)^{n-j} (1 - j/t)\right\}. \qquad (*)$$

Using Bernoulli's inequality $1 - j/t \leqslant (1 - 1/t)^j$ for $j = 0, 1, \ldots, t - 1$, we obtain

$$E(\min\{X,t\}) \geqslant t\left\{1 - \sum_{j=0}^{t-1} \binom{n}{j} (p - p/t)^j (1-p)^{n-j}\right\}$$

$$\geqslant t\left\{1 - \sum_{j=0}^{n} \binom{n}{j} (p - p/t)^j (1-p)^{n-j}\right\}$$

$$= t\{1 - (1 - p/t)^n\}.$$

Editorial comment. Let

$$Z = \min\{X, t\} = (X + t)/2 - |Y|/2,$$

where $Y = X - t$. Thus

$$E(Z) = (pn + t)/2 - E(|Y|)/2.$$

Hence lower and upper bounds for $E(Z)$ can be obtained by finding upper and lower bounds, respectively, for $E(|Y|)$.

Since the square-root function is convex, Jensen's inequality gives

$$E(|Y|) = E\left(\sqrt{Y^2}\right) \leqslant \sqrt{E(Y^2)}\,.$$

Now

$$E(Y^2) = E(X - t)^2 = E(X - np + np - t)^2$$

$$= E(X - np)^2 + 2(np - t)E(X - np) + (np - t)^2$$

$$= npq + (np - t)^2, \qquad (**)$$

where $q = 1 - p$. Thus a lower bound for $E(Z)$ different from that given in (a) is

$$E(Z) \geqslant (pn + t)/2 - \sqrt{npq + (np - t)^2}\,/2.$$

If t is much less or much greater than np, rather accurate upper bounds can be given for $E(Z)$ by using $(*)$. When t is close to np, we can proceed as follows.

Hölder's inequality gives

$$E(Y^2) = E\left(|Y|^{2/3} \cdot |Y|^{4/3}\right) \leqslant \left\{E(|Y|)\right\}^{2/3}\left\{E(Y^4)\right\}^{1/3}$$

or

$$E(|Y|) \geqslant \left\{E(Y^2)\right\}^{3/2}\left\{E(Y^4)\right\}^{-1/2}.$$

Thus we have the upper bound

$$E(Z) \leqslant (pn + t)/2 - \left\{E(Y^2)\right\}^{3/2}\left\{E(Y^4)\right\}^{-1/2}/2,$$

where $E(Y^2)$ is given by $(**)$ and

$$\begin{aligned} E(Y^4) &= E(X - t)^4 = E(X - np + np - t)^4 \\ &= E(X - np)^4 + 4(np - t)E(X - np)^3 + 6(np - t)^2E(X - np)^2 \\ &\quad + 4(np - t)^3E(X - np) + (np - t)^4 \\ &= npq(1 - 6pq) + 3(npq)^2 + 4(np - t)npq(q - p) \\ &\quad + 6(np - t)^2npq + (np - t)^4. \end{aligned}$$

Part (a) was solved also by R. A. Agnew, F. T. Bruss, D. Callan, J. Guillerme (France), R. J. Hendel, F. Levron (France), S. Philipp, H. L. Stubbs, and the Chico Problem Group. Several of these solvers also included upper bounds.

Zeros of Some Special Polynomials

E 3348 [1989, 735]. *Proposed by Alfred Witkowski, University of Wroclaw, Poland.*

(a) Suppose k is a given nonnegative integer. If $n > k$ and

$$P_n(x) = \prod_{j=0}^{n}(x - j),$$

let $x_{k,n}$ be the zero of P_n' in the interval $(k, k + 1)$. Find $\lim_{n\to\infty} x_{k,n}$ and estimate the speed of convergence.

(b) If $n > 0$ and

$$Q_n(x) = \prod_{j=0}^{n}(x - j^2),$$

let t_n be the zero of Q_n' in the interval $(0, 1)$. Find $\lim_{n\to\infty} t_n$ and estimate the speed of convergence.

Solution by Richard Stong, University of California at Los Angeles. The zero $x_{k,n}$ converges to k as $n \to \infty$; more precisely $x_{k,n} = k + (\log n)^{-1} + O((\log n)^{-2})$. The zero t_n converges to the unique solution t in $(0, 1)$ of $\pi\sqrt{t}\,\cot(\pi\sqrt{t}) = -1$, roughly $t = 0.4170236\ldots$; more precisely $t_n = t + cn^{-1} + O(n^{-2})$, where $c = 4t^2/(\pi^2 t + 2) = 0.11374\ldots$.

Note that $P_n'(x) = P_n(x)f_n(x)$, where

$$f_n(x) = \sum_{j=0}^{n}(x - j)^{-1}.$$

Finally

$$\begin{aligned} h'(t) &= -\frac{\pi^2}{4t}\csc^2(\pi\sqrt{t}) - \frac{\pi}{4t^{3/2}}\cot(\pi\sqrt{t}) - \frac{1}{2t^2} \\ &= -\frac{\pi^2}{4t}\left(1 + \frac{1}{\pi^2 t}\right) + \frac{\pi}{4t^{3/2}}\frac{1}{\pi t^{1/2}} - \frac{1}{2t^2} \\ &= -\frac{\pi^2}{4t} - \frac{1}{2t^2}, \end{aligned}$$

so that $-1/h'(t) = 4t^2/(\pi^2 t + 2)$. Thus the result claimed is established.

Solved also by W. E. Briggs, W. Janous (Austria), O. P. Lossers (The Netherlands), R. Martin, L. F. Martins, J.-M. Monier (France), K. Zacharias, The NSA Problems Group, the Western Maryland College Problems Group, and the proposer.

A Sequence Defined by the Limits of Integration

E 3383 [1990, 343]. *Proposed by Harold G. Diamond, University of Illinois, Urbana, IL.*

Suppose α is a given number in $(0, 1]$. Show that there exists a unique sequence of positive numbers $a_0 < a_1 < a_2 < \cdots$ such that $a_n/n \to 1$ as $n \to \infty$ and

$$\int_{a_n}^{a_{n+1}} \frac{1 + \cos t}{t}\, dt = \frac{1}{n + \alpha} \qquad (n = 0, 1, 2, \ldots).$$

Solution by David S. Ross, Kodak Research Labs, Rochester, NY. For any $a_0 > 0$, the integral $\int_{a_0}^{x}(1 + \cos t)t^{-1}\, dt$ is a positive, strictly increasing function of x on $[a_0, \infty)$ that approaches ∞ as $x \to \infty$; its integrand is positive except at isolated points, and the integral is the sum of $\ln(x/a_0)$ and a truncation of the convergent integral $\int_{a_0}^{\infty}(\cos t)t^{-1}\, dt$. Thus, any choice of $a_0 > 0$ uniquely determines a strictly increasing sequence $\{a_n\}_{n=1}^{\infty}$ for which the integrals have the desired values.

To determine the growth rate of a_n, note first that $a_n \to \infty$ as $n \to \infty$, since $\int_{a_0}^{a_{n+1}}(1 + \cos t)t^{-1}\, dt = \sum_{i=0}^{n}(i + \alpha)^{-1}$, which goes to ∞ as $n \to \infty$. Now define $\gamma_\alpha = \lim_{n\to\infty}\{\sum_{i=0}^{n}(i + \alpha)^{-1} - \ln n\}$. The existence of this limit is proved in much the same way as the existence of Euler's constant γ. The condition $a_n/n \to 1$ is equivalent to $\lim_{n\to\infty}(\ln a_n - \ln n) = 0$. Since

$$\ln a_n - \ln n = \ln a_0 - \int_{a_0}^{a_n} \frac{\cos t}{t}\, dt + \sum_{i=0}^{n-1} \frac{1}{i + \alpha} - \ln n,$$

we have

$$\lim_{n\to\infty} (\ln a_n - \ln n) = \ln a_0 - \int_{a_0}^{\infty} \frac{\cos t}{t}\, dt + \gamma_\alpha.$$

As a function of a_0, the right side of this equation is strictly increasing for $a_0 > 0$ and ranges from $-\infty$ to ∞. Hence there is a unique value of $a_0 > 0$ for which the right side of this equation is 0. This value of a_0 determines the unique sequence with the desired properties.

Editorial comment. Needless to say, the assertion of the problem is valid for any positive α, since increasing α by 1 merely replaces a_n by a_{n+1} $(n = 0, 1, 2, \ldots)$. Also the condition $a_n/n \to 1$ can obviously be generalized to $a_n/n \to \beta$ for any fixed positive number β. Further generalizations are possible.

Solved also by M. Golomb, B. Haible (Germany), E. A. Herman, K. W. Lau (Hong Kong), E. Levine, J. H. Lindsey II, O. P. Lossers (The Netherlands), J. G. Merickel (student), J.-M. Monier (France), W. A. Newcomb, T. S. Norfolk, A. Riese, D. B. Tyler, P. J. Zwier, University of South Alabama Problem Group, Western Maryland College Problems Group, and the proposer.

A Perfect Property of Handful of 2 Numbers

E 3385 [1990, 427]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles.*

For what positive integers n is it true that

$$\sum_{p \leqslant \pi(n)} p = n.$$

Here, as usual, $\pi(x)$ denotes the number of primes not exceeding x and the summation is over primes p.

Solution by Marcin E. Kuczma, University of Warsaw, Poland. The answer is: $5, 17, 41, 77, 100$.

Let p_k be the k-th prime number, so that $p_1 = 2$, $p_2 = 3, \ldots$. Define $n_k = p_1 + p_2 + \cdots + p_k$. A positive integer n satisfying the condition of the problem must lie in the sequence $\{n_k\}_{k=1}^{\infty}$ and moreover must satisfy

$$\sum_{p \leqslant \pi(n_k)} p = n_k.$$

Thus p_k must be the largest prime not exceeding $\pi(n_k)$ and so

$$\pi(\pi(n_k)) = k. \tag{1}$$

Thus the problem may be restated as that of finding all k satisfying (1); the corresponding values of n_k provide the solution of the problem as stated.

An easy induction using $p_k \geqslant 2k - 1$ shows that

$$n_k > k^2 \quad \text{for } k = 1, 2, \ldots. \tag{2}$$

Less trivial is the inequality

$$\pi(n) > n^{1/\sqrt{2}} \quad \text{for } n \geqslant 149. \tag{3}$$

Postponing the proof of (3) for the moment, we note that it implies

$$\pi(\pi(n)) > n^{1/2} \quad \text{for } n \geqslant 859, \tag{4}$$

because $\pi(859) = 149$. Since $n_{23} = 874 > 859$, relations (2) and (4) imply

$$\pi(\pi(n_k)) > k \quad \text{for } k \geqslant 23.$$

Among the numbers between 1 and 22 inclusive, only $k = 2, 4, 6, 8, 9$ satisfy (1); accordingly $n_k = 5, 17, 41, 77, 100$ constitute the answer.

To justify (3) we use the estimate ([1], Theorem 2)

$$\pi(n) > \frac{n}{\ln n - 1/2} \quad \text{for } n \geqslant 67, \tag{5}$$

together with

$$\frac{\ln(\ln n - 1/2)}{\ln n} < 1 - 1/\sqrt{2} \quad \text{for } n \geqslant 241. \tag{6}$$

(The inequality (6) holds for $n = 241$ and hence for all $n \geqslant 241$, since $(\ln x)/(x + 1/2)$ decreases for $x \geqslant 4$.) From (5) and (6) we obtain (3) for $n \geqslant 241$. For $149 \leqslant n \leqslant 241$ it is easy to verify (3) directly.

We remark that (5) is a rather advanced tool. To make the argument fully elementary, we might apply instead the Chebyshev-type estimate

$$\pi(n) > (2n)/(3 \ln n) \quad \text{for } n \geqslant 200, \tag{7}$$

which is provable easily from the factorization of binomial coefficients (cf. [2]). The inequality (7) implies (3) for $n \geqslant 6737$, while the range from 149 to 6736 must be examined computationally.

REFERENCES

1. J. Barkley Rosser and Lowell Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.*, 6 (1962) 64–94.
2. Don Bernard Zagier, Die ersten 50 Millionen Primzahlen, *Elemente der Mathematik*, Supplement No. 15, Birkhauser, 1977. (English translation in *Math. Intelligencer* 0 (1977), 7–19.)

Solved also by B. M. M. der Weger (Netherlands), Kevin Ford, L. L. Foster, J. Herzog & P. R. Smith (Germany), J. McHugh, A. Pedersen (Denmark), R. Stong, and the proposer. Five incorrect or incomplete solutions were received.

Creating a Polynomial from the Distance to the Nearest Integer

E 3389 [1990, 428]. *Proposed by Joseph H. Silverman, Brown University, Providence, RI.*

Let $\|t\|$ be the distance from the real number t to the nearest integer. Prove that there is a unique real number r greater than 1 such that

$$\sum_{n=1}^{\infty} r^{-2n} \|r^{n-1} t\|$$

is a polynomial in $\|t\|$; give this polynomial explicitly for the special value of r referred to.

Solution by Marcin E. Kuczma, University of Warsaw, Warsaw, Poland. The desired value is $r = 2$, which produces the polynomial $(1/2)(\|t\| - \|t\|^2)$.

Let $f_r(t) = \sum_{n=1}^{\infty} r^{-2n} \|r^{n-1} t\|$. For $r > 1$, f_r is a continuous, nonnegative function. If r is also an integer, then f_r is an even function, periodic with period 1. Hence $f_r(1 - t) = f_r(t)$ if r is any integer at least 2. For any real $r > 1$, if we isolate the first term in the series and restrict attention to the interval $[0, 1/2]$, we obtain the functional equation

$$r^2 f_r(t) = t + f_r(rt), \qquad t \in [0, 1/2]. \tag{1}$$

Assume that f_r is a polynomial in $\|t\|$. In particular, $f_r(t)$ depends only on $\|t\|$ and so $f_r(1) = f_r(0) = 0$. Hence all terms of the series vanish for $t = 1$. The vanishing of the second term yields $\|r\| = 0$, so r is an integer at least 2. Since f_r restricted to $[0, 1/2]$ is a polynomial, both sides of (1) are polynomials on the interval $[0, 1/(2r)]$. Differentiating twice yields $f_r''(t) = f_r''(rt)$ on $[0, 1/(2r)]$, showing that f_r'' is constant on that interval. Hence the polynomial in question has degree at most 2. Since f_r is nonnegative, the functional equation implies that $r^2 f_r(t) - t \geqslant 0$ for $t \in [0, 1/2]$, with equality for $t = 0$ and $t = 1/r$ (this uses $r \geqslant 2$). We now have a quadratic function vanishing at 0 and $1/r$ and nonnegative throughout $[0, 1/2]$. This requires that $1/r$ is the endpoint of the interval, i.e, $r = 2$.

To prove the formula claimed for $f_2(t)$, it suffices to show that the continuous function $g(t) = f_2(t) - t(1 - t)/2$ is 0 throughout $[0, 1]$. The symmetry of f_2 about $1/2$ implies $g(t) = g(1 - t)$ for $t \in [0, 1]$, and the functional equation implies $4g(t) = g(2t)$ for $t \in [0, 1/2]$. These two relationships imply $\max_{[0, 1/2]} g = \max_{[1/2, 1]} g = \max_{[0, 1]} g = 4 \max_{[0, 1/2]} g$. Hence $\max g = 0$, and similarly $\min g = 0$.

Solved also by O. P. Lossers (Netherlands), T. S. Norfolk, the late David R. Richman, R. Stong, A. Tissier (France), A. Zulauf (New Zealand), and the proposer.

The Golden Ratio Revisited

E 3391 [1990, 528]. *Proposed by David Doster, Choate Rosemary Hall, Wallingford, CT.*

Define a function on the positive integers by putting $f(1) = 1$ and setting

$$f(n + 1) = \begin{cases} f(n) + 2 & \text{if } f(f(n) - n + 1) = n \\ f(n) + 1 & \text{if otherwise.} \end{cases}$$

(i) Find and prove a simple explicit formula for $f(n)$ involving the greatest integer function.

(ii) Show that if $f(f(n) - n + 1) \neq n$, then $f(f(n) - n + 1) = n + 1$.

Solution by Nathan J. Fine, Deerfield Beach, FL. The formula is $f(n) = \lfloor \phi n \rfloor$, where $\phi = (1 + \sqrt{5})/2$ is the golden ratio, satisfying $\phi^2 - \phi = 1$. We prove this inductively by showing that $F(n) = \lfloor \phi n \rfloor$ satisfies the same recurrence as $f(n)$ (note that $F(1) = 1$). This will include a proof that F satisfies (ii).

First we observe that $F(n + 1) - F(n)$ is 1 or 2. This follows from $F(n + 1) - F(n) < \phi(n + 1) - (\phi n - 1) = \phi + 1 < 3$ and $F(n + 1) - F(n) > \phi(n + 1) - 1 - \phi n = \phi - 1 > 0$.

Next, let $H(n) = F(F(n) - n + 1)$. The fact that $H(n)$ is n or $n + 1$ follows from $H(n) < \phi(\phi n - n + 1) = (\phi^2 - \phi)n + \phi < n + 2$ and $H(n) > \phi(\phi n - 1 - n + 1) - 1 = (\phi^2 - \phi)n - 1 = n - 1$.

Now suppose $F(n + 1) - F(n) = 1$. This implies

$$H(n) = F(F(n + 1) - n) = \lfloor \phi(\lfloor \phi(n + 1) \rfloor - n) \rfloor > \phi(\phi(n + 1) - 1 - n) - 1$$
$$= \phi(\phi - 1)n + \phi(\phi - 1) - 1 = n,$$

so we conclude that $H(n) = n + 1$.

On the other hand, $F(n+1) - F(n) = 2$ implies $H(n) = F(F(n+1) - n - 1) < \phi(\phi(n+1) - n - 1) = \phi(\phi - 1)n + \phi(\phi - 1) = n + 1$, so this requires $H(n) = n$. This completes the inductive proof that $F = f$.

Solved also by the proposer and 39 others.

A Simple Sequence of Complex Numbers

E 3396 [1990, 529]. *Proposed by Bruno Haible (student), Karlsruhe, Germany.*

Suppose we define a sequence $\{x_n\}_{n=0}^{\infty}$ by putting $x_0 = 1$, $x_1 = 1/2$, and

$$(n^2+1)x_n = (2n^2 - n)x_{n-1} - (n^2 - n)x_{n-2} \quad \text{for } n = 2, 3, \ldots .$$

Prove that the sequence $\{x_n\}$ is bounded and determine $\liminf x_n$ and $\limsup x_n$.

Composite solution, based on the solutions of Jean Anglesio, Garches, France, and Richard Stong, University of California, Los Angeles. We show that $\liminf x_n = -\sqrt{\pi/\sinh \pi}$ and $\limsup x_n = \sqrt{\pi/\sinh \pi}$.

Define a sequence $\{y_n\}_{n=0}^{\infty}$ by $y_0 = 0$ and

$$y_n = n(x_n - x_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{1}$$

Then the defining recurrence for x_n, together with the value of x_1, yields

$$x_n = -n(y_n - y_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{2}$$

Now let $z_n = x_n + iy_n$, where $i = \sqrt{-1}$. Then $z_0 = 1$, and (1) and (2) yield

$$z_n = in(z_n - z_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{3}$$

Solving (3) for z_n we obtain $z_n = (1 + i/n)^{-1} z_{n-1}$, so

$$z_n = \prod_{k=1}^{n} \left(1 + \frac{i}{k}\right)^{-1}.$$

Now let $(1 + i/k)^{-1} = r_k e^{-i\theta_k}$, where $r_k > 0$ and $0 < \theta_k < \pi/2$. Then $r_k = \sqrt{k^2/(k^2+1)}$ and $\theta_k = \tan^{-1}(1/k)$. Thus

$$z_n = \left(\prod_{k=1}^{n} \frac{k^2}{k^2+1}\right)^{1/2} e^{-i(\theta_1 + \cdots + \theta_n)},$$

so

$$x_n = \operatorname{Re} z_n = \left(\prod_{k=1}^{n} \frac{k^2}{k^2+1}\right)^{1/2} \cos(\theta_1 + \cdots + \theta_n).$$

Now from the infinite product $\prod_{k=1}^{\infty}(1 + u^2/k^2) = (\sinh \pi u)/(\pi u)$, we obtain

$$\left(\prod_{k=1}^{\infty} \frac{k^2}{k^2+1}\right)^{1/2} = \sqrt{\frac{\pi}{\sinh \pi}}.$$

Since $\theta_k = \tan^{-1}(1/k) = k^{-1} + O(k^{-3})$, we see that $\theta_1 + \theta_2 + \cdots + \theta_n$ comes arbitrarily close to both even and odd multiples of π, and the conclusion follows.

Solved also by S. Chen, C. Georghiou (Greece), I. Kastanas, O. P. Lossers (The Netherlands), F. W. Schmidt, J. S. Sumner, the University of South Alabama Problem Group, and the proposer.

On the other hand, $F(n+1) - F(n) = 2$ implies $H(n) = F(F(n+1) - n - 1) < \phi(\phi(n+1) - n - 1) = \phi(\phi - 1)n + \phi(\phi - 1) = n + 1$, so this requires $H(n) = n$. This completes the inductive proof that $F = f$.

Solved also by the proposer and 39 others.

A Simple Sequence of Complex Numbers

E 3396 [1990, 529]. *Proposed by Bruno Haible (student), Karlsruhe, Germany.*

Suppose we define a sequence $\{x_n\}_{n=0}^{\infty}$ by putting $x_0 = 1$, $x_1 = 1/2$, and

$$(n^2+1)x_n = (2n^2 - n)x_{n-1} - (n^2 - n)x_{n-2} \quad \text{for } n = 2, 3, \ldots .$$

Prove that the sequence $\{x_n\}$ is bounded and determine $\liminf x_n$ and $\limsup x_n$.

Composite solution, based on the solutions of Jean Anglesio, Garches, France, and Richard Stong, University of California, Los Angeles. We show that $\liminf x_n = -\sqrt{\pi/\sinh \pi}$ and $\limsup x_n = \sqrt{\pi/\sinh \pi}$.

Define a sequence $\{y_n\}_{n=0}^{\infty}$ by $y_0 = 0$ and

$$y_n = n(x_n - x_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{1}$$

Then the defining recurrence for x_n, together with the value of x_1, yields

$$x_n = -n(y_n - y_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{2}$$

Now let $z_n = x_n + iy_n$, where $i = \sqrt{-1}$. Then $z_0 = 1$, and (1) and (2) yield

$$z_n = in(z_n - z_{n-1}) \quad \text{for } n = 1, 2, \ldots . \tag{3}$$

Solving (3) for z_n we obtain $z_n = (1 + i/n)^{-1} z_{n-1}$, so

$$z_n = \prod_{k=1}^{n} \left(1 + \frac{i}{k}\right)^{-1}.$$

Now let $(1 + i/k)^{-1} = r_k e^{-i\theta_k}$, where $r_k > 0$ and $0 < \theta_k < \pi/2$. Then $r_k = \sqrt{k^2/(k^2+1)}$ and $\theta_k = \tan^{-1}(1/k)$. Thus

$$z_n = \left(\prod_{k=1}^{n} \frac{k^2}{k^2+1}\right)^{1/2} e^{-i(\theta_1 + \cdots + \theta_n)},$$

so

$$x_n = \operatorname{Re} z_n = \left(\prod_{k=1}^{n} \frac{k^2}{k^2+1}\right)^{1/2} \cos(\theta_1 + \cdots + \theta_n).$$

Now from the infinite product $\prod_{k=1}^{\infty}(1 + u^2/k^2) = (\sinh \pi u)/(\pi u)$, we obtain

$$\left(\prod_{k=1}^{\infty} \frac{k^2}{k^2+1}\right)^{1/2} = \sqrt{\frac{\pi}{\sinh \pi}}.$$

Since $\theta_k = \tan^{-1}(1/k) = k^{-1} + O(k^{-3})$, we see that $\theta_1 + \theta_2 + \cdots + \theta_n$ comes arbitrarily close to both even and odd multiples of π, and the conclusion follows.

Solved also by S. Chen, C. Georghiou (Greece), I. Kastanas, O. P. Lossers (The Netherlands), F. W. Schmidt, J. S. Sumner, the University of South Alabama Problem Group, and the proposer.

ADVANCED PROBLEMS

6670. *Proposed by R. H. Jeurissen, Toernooiveld, Nijmegen, The Netherlands.*

Let $\{0, 1\}^n$ denote the set of n-bit strings of zeros and ones. If $(a_1, a_2, \ldots, a_n) \in \{0, 1\}^n$, let $\pi_n(a_1, a_2, \ldots, a_n)$ be the string $(b_1, b_2, \ldots, b_n)$ given by $b_1 = a_1$ and $b_k \equiv a_k + a_{k-1} \pmod 2$ for $1 < k \leqslant n$. Since $(a_1, a_2, \ldots, a_n)$ can be retrieved from $(b_1, b_2, \ldots, b_n)$, it is clear that π_n is a permutation of $\{0, 1\}^n$. Determine the cycle structure of the permutation π_n, i.e., the lengths of the cycles that occur and the number of cycles of each length.

6671. *Proposed by Carl Pomerance, University of Georgia, Athens, GA.*

Let $V(x)$ denote the number of distinct values not exceeding x taken on by Euler's arithmetical function ϕ. Let $V^*(x)$ denote the number of these values with a unique pre-image. For example, $V(15) = 7$, $V^*(15) = 0$.

R. D. Carmichael conjectured that $V^*(x) = 0$ for all x. Prove the weaker assertion that $\liminf_{x \to \infty}\{V^*(x)/V(x)\} < 1$.

6672. *Proposed by H. B. Kushner, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY.*

If a and b are positive real numbers, prove that

$$\int_0^{\pi/2}\{(a\cos^2\theta + b\sin^2\theta)(a\sin^2\theta + b\cos^2\theta)\}^{-1/2}\,d\theta$$

$$= \int_0^{\pi/2}\{a^2\cos^2\theta + b^2\sin^2\theta\}^{-1/2}\,d\theta$$

and use it to prove that the integral on the right is unchanged if a and b are replaced by $(ab)^{1/2}$ and $(a + b)/2$, respectively.

SOLUTIONS OF ADVANCED PROBLEMS

A Refinement of Descartes' Rule of Signs

6597 [1989, 265]. *Proposed by Ichiro Murase, Science University of Tokyo, Japan.*

Let $f(x) \in \mathbb{R}[x]$, say

$$f(x) = a_0 + a_1x + \cdots + a_nx^n, \qquad n \geqslant 0, \qquad a_n \neq 0.$$

Let $Z_+(f)$ denote the number of zeros of f in $(0, +\infty)$, each counted according to its multiplicity. Let $W(a_0, a_1, \ldots, a_n)$ denote the number of pairs (a_k, a_l) with $k < l$ such that $a_ka_l < 0$ and $a_v = 0$ for $k < v < l$. Descartes' rule of signs says that $Z_+(f) \leqslant W(a_0, a_1, \ldots, a_n)$ and $W(a_0, a_1, \ldots, a_n) - Z_+(f)$ is even.

If there is an index k, $1 \leqslant k \leqslant n - 1$, such that

$$a_{k-1}a_k < 0, \qquad a_ka_{k+1} < 0, \qquad a_k^2 \leqslant a_{k-1}a_{k+1},$$

prove that

$$Z_+(f) \leqslant W(a_0, a_1, \ldots, a_n) - 2.$$

Solution by Allan Pedersen, Søborg, Denmark. Introduce the notation $W(f) = W(a_0, a_1, \ldots, a_n)$ and define $f^*(x) = x^n f(1/x) = a_n + a_{n-1}x + \cdots + a_0 x^n$.

LEMMA 1. *$W(f^*) = W(f)$ and $Z_+(f^*) = Z_+(f)$.*

Proof. The first assertion is clear. For the second, note that the statement $\alpha > 0$ is a root of $f(x)$ of multiplicity k means $f(x) = (x - \alpha)^k g(x)$, $g(\alpha) \neq 0$. But then

$$f^*(x) = (\alpha^{-1} - x)^k \alpha^k x^{n-k} g(1/x),$$

so $\alpha^{-1} > 0$ is a root of f^* of multiplicity k. Since $f^{**}(x) = f(x)$, the result easily follows.

LEMMA 2. *If $Z_+(f) = W(f)$, then $Z_+(f') = W(f')$.*

Proof. Since $f'(x) = a_1 + 2a_2 x + \cdots + na_n x^{n-1}$, either

$$W(f') = W(f) \quad \text{or} \quad W(f') = W(f) - 1.$$

By a well-known argument using Rolle's Theorem

$$Z_+(f') \geqslant Z_+(f) - 1.$$

Hence

$$Z_+(f') \geqslant W(f') - 1 \quad \text{or} \quad Z_+(f') \geqslant W(f').$$

Since $W(f') - Z_+(f') \geqslant 0$ is even, the last alternative must hold.

LEMMA 3. *Let $f(x) = \sum_{j=0}^{n} \binom{n}{j} c_j x^j$ where $n \geqslant 2$, the c_j are real, and $c_n \neq 0$, and set*

$$g_k(x) = c_{k-1} + 2c_k x + c_{k+1} x^2, \qquad 1 \leqslant k \leqslant n - 1.$$

If $Z_+(f) = W(f)$ and $g_k(x)$ is not identically zero, then $Z_+(g_k) = W(g_k)$.

Proof. This is clear for $n = 2$; proceed by induction on n. For $n > 2$,

$$f'(x) = n \sum_{j=0}^{n-1} \binom{n-1}{j} c_{j+1} x^j$$

is a polynomial of the type considered. By Lemma 2,

$$Z_+(f'/n) = W(f'/n).$$

Hence by induction $Z_+(g_k) = W(g_k)$ for $2 \leqslant k \leqslant n - 1$. However, c_0 does not occur in $f'(x)$, so the induction will not be completed until the result for $k = 1$ is established. Consider

$$f^*(x) = \sum_{j=0}^{n} \binom{n}{j} c_{n-j} x^j = \sum_{j=0}^{n} \binom{n}{j} c_j^* x^j,$$

where $c_j^* = c_{n-j}$. By Lemma 1, $Z_+(f^*) = W(f^*)$. Hence by the case $k = n - 1$ we have

$$Z_+(h) = W(h),$$

where

$$h(x) = c_{n-2}^* + 2c_{n-1}^* x + c_n^* x^2 = c_2 + 2c_1 x + c_0 x^2 = g_1^*(x).$$

Hence $Z_+(g_1) = W(g_1)$ and Lemma 3 is proved.

Now return to the original problem, and introduce numbers c_j such that

$$f(x) = c_0 + \binom{n}{1} c_1 x + \cdots + \binom{n}{k} c_k x^k + \cdots + c_n x^n.$$

Since $c_k \neq 0$, the hypotheses on the a_j yield

$$c_k^2 < \frac{k+1}{k} \cdot \frac{n-k+1}{n-k} c_k^2 \leqslant c_{k-1} c_{k+1}.$$

If $Z_+(f) = W(f)$ then Lemma 3 yields $Z_+(g_k) = W(g_k)$ where $g_k(x) = c_{k-1} + 2c_k x + c_{k+1} x^2$. The hypotheses on the a_j now yield

$$2 = W(g_k) = Z_+(g_k).$$

Hence, $4c_k^2 - 4c_{k-1}c_{k+1} \geqslant 0$, a contradiction. The result of the problem follows.

Editorial comment. The solutions of John Steinig and the proposer did not use derivatives. They noted that if the sequence of coefficients of $f(x)$ has ω changes of sign, and if $\beta > 0$, then the sequence of coefficients of $g(x) = (x + \beta)f(x)$ has at most ω changes of sign (see Pólya-Szegö, II.V.32). Thus if β can be chosen so the coefficients of $g(x)$ have at most $\omega - 1$ changes of sign, Descartes' rule applied to $g(x)$ yields

$$Z_+(f) = Z_+(g) \leqslant \omega - 1$$

and hence (by its parity assertion) $Z_+(f) \leqslant \omega - 2$. It suffices to establish the existence of such a number β, and this they did.

Steinig carried the analysis somewhat further. He showed that if f has N coefficients $a_{k_1}, \ldots, a_{k_N}$ with $k_{i-1} \leqslant k_i - 2$ $(i = 2, \ldots, N)$ such that

$$a_{v-1}a_v < 0, \qquad a_v a_{v+1} < 0, \quad \text{and} \quad a_v^2 \leqslant a_{v-1}a_{v+1}$$

for $v = k_1, \ldots, k_N$, and if the intervals

$$\left[-\frac{a_v}{a_{v+1}}, -\frac{a_{v-1}}{a_v} \right], \qquad v = k_1, \ldots, k_N$$

have a nonempty intersection, then

$$Z_+(f) \leqslant W(a_0, \ldots, a_n) - 2N.$$

Moreover, he exhibited polynomials of degree n for each n such that equality holds. He also observed that if $k < l < m$ and a_k, a_l, a_m are consecutive coefficients of $f(x)$ but k, l, m are not consecutive integers (i.e., $m - k > 2$ and $a_v = 0$ for $k < v < m$, $v \neq l$) the conditions

$$a_k a_l < 0, \qquad a_l a_m < 0, \qquad a_l^2 < a_k a_m$$

do *not*, in general, imply

$$Z_+(f) \leqslant W(a_0, \ldots, a_n) - 2.$$

For example, $W(1, -7, 50) = 2$, but for $f(x) = x^6 - 7x^4 + 50$ we have

$$f(2) > 0, \qquad f\left(\sqrt{14/3}\right) < 0, \quad \text{and} \quad f(3) > 0.$$

Finally, Steinig remarks that this problem is related to results of Fekete and Pólya (*Rend. Circolo Mat. Palermo* 34 (1912), 1–32), and perhaps even more closely to a result of E. Bálint mentioned towards the end of that paper.

Solved also by J. Steinig (Switzerland), S. Gaignoux (France), and the proposer.

A Congruence Related to an Elliptic Curve

6604 [1989, 533]. *Proposed by Michael Reid, University of California, Berkeley.*

For which primes $p > 3$ is it true that

$$\sum_{k=0}^{p-1} \frac{(3k)!}{(k!)^3}(-6)^{3p-3k} \equiv 0 \pmod{p}?$$

Combined Solution by Noam D. Elkies, Harvard University, Cambridge, MA and Ronald J. Evans, University of California, San Diego. The answer is: the primes congruent to 2 modulo 3. More specifically, we prove the following theorem.

THEOREM. *For each prime $p > 3$, define*

$$S = \sum_{k=0}^{p-1} \frac{(3k)!}{k!^3}(-6)^{3p-3k}.$$

Then

$$S \equiv \begin{cases} 0 \pmod{p}, & \text{if } p \equiv 2 \pmod{3}, \\ 216\left(\frac{p-1}{3}\right)!^{-3} \pmod{p}, & \text{if } p \equiv 1 \pmod{3}. \end{cases}$$

Proof. All congruences below are (mod p). Let $r = \lfloor p/3 \rfloor$. Then $S \equiv -216T$, where

$$T := \sum_{k=0}^{r} \frac{(p-1)!6^{p-3k-1}}{(p-3k-1)!k!^3},$$

so we must show that

$$T \equiv \begin{cases} 0, & \text{if } p = 3r+2, \\ -1/r!^3, & \text{if } p = 3r+1. \end{cases}$$

Since

$$(u^3+v^3+w^3+6uvw)^{p-1} = \sum_{\substack{a,b,c,d \geqslant 0 \\ a+b+c+d=p-1}} \frac{(p-1)!6^d}{a!b!c!d!} u^{3a+d}v^{3b+d}w^{3c+d},$$

T is the coefficient of $(uvw)^{p-1}$ in the expansion of $(u^3+v^3+w^3+6uvw)^{p-1}$. Since $\sum_{u \pmod p} u^k \equiv 0$ for $0 \leqslant k \leqslant p-2$,

$$\sum_{u,v,w \pmod p} (u^3+v^3+w^3+6uvw)^{p-1} \equiv T \sum_{u,v,w \pmod p} (uvw)^{p-1}$$

$$\equiv T\left(\sum_{u \pmod p} u^{p-1}\right)^3 \equiv -T.$$

Thus,

$$T \equiv \sum_{u,v,w(\mathrm{mod}\ p)} \left\{1 - (u^3 + v^3 + w^3 + 6uvw)^{p-1}\right\},$$

so $T \equiv N$, where N is the number of $(u, v, w) \in \mathbb{F}_p^3$ on the cubic $u^3 + v^3 + w^3 + 6uvw = 0$ (i.e., 1 more than $(p-1)$ times the number of $\mathbb{F}_p$-rational points on the projective curve defined by the same cubic equation). From the transformation

$$u = -2x + y + z, \qquad v = -2x - y + z, \qquad w = -2x - 2z$$

(of determinant -12), we see that N equals the number of $(x, y, z) \in \mathbb{F}_p^3$ on $zy^2 = 4x^3 + z^3$. If $p = 3r + 2$, then $x \to x^3$ is a bijection of the field of p elements, so $N = p^2$ and $T \equiv N \equiv 0$. Finally, suppose that $p = 3r + 1$. Then there exists a primitive cube root of unity ρ (mod p). Since

$$3(x^3 + y^3 + z^3 + 6xyz) = (x + y + z)^3 + (x + \rho y + \rho^2 z)^3 + (x + \rho^2 y + \rho z)^3,$$

we have $N = M$, where M is the number of $(u, v, w) \in \mathbb{F}_p^3$ on the Fermat cubic $u^3 + v^3 + w^3 = 0$. Reasoning as above with $u^3 + v^3 + w^3$ in place of $u^3 + v^3 + w^3 + 6uvw$, we find that M is congruent (mod p) to the coefficient of $(uvw)^{p-1}$ in the expansion of $(u^3 + v^3 + w^3)^{p-1}$, i.e., $M \equiv (3r)!/(r!)^3$. Thus $T \equiv N = M \equiv -1/r!^3$.

Remark. It was already known to Jacobi that $-1/r!^3 \equiv A$ (mod p) for primes $p = 3r + 1$, where A is uniquely determined by $4p = A^2 + 27B^2, 3|(A + 1)\ A \in \mathbb{Z}, B \in \mathbb{Z}$.

Comments by N. D. Elkies. More generally we may consider the sum

$$S_a(p) := \sum_{k=0}^{p-1} \binom{3k}{k, k, k} a^{-k} \bmod p$$

for nonzero rational a, corresponding to the curve

$$X^3 + Y^3 - a^{-1}Z^3 + XYZ \equiv 0 \bmod p.$$

This curve becomes singular when $a = 27$, when it factors into the product of the line $Z = 3(X + Y)$ and the conic

$$Z^2 + 3Z(X + Y) + 9(X^2 - XY + Y^2) = 0;$$

their two points of intersection are defined over $\mathbb{F}_p$ if and only if $(-3/p) = 1$, so their total number of points mod p is $3p^2 - 3p + 1$ or $p^2 + p - 1$ for $p \equiv 1$ or $p \equiv 2$ (mod 3), respectively, whence $S_{27}(p) = +1$ in the former case and -1 in the latter case. For $a \neq 27$ that cubic is an elliptic curve of j-invariant $(a + 216)^3 a/(a - 27)^3$. Thus for $a = -216$, 54 or 24, the j-invariant 0, 54000 or -12288000 indicates complex multiplication by the orders $\mathbb{Z}[\rho]$, $\mathbb{Z}[\sqrt{-3}]$, or $\mathbb{Z}[3\rho]$ in $\mathbb{Q}[\sqrt{-3}]$, where ρ is a complex primitive cube root of unity, and so $S_a(p) = 0$ iff $p \equiv -1$ (mod 3), as in the solution above; whereas for all other values of a the elliptic curve has no complex multiplication, and then all that can be said is that $S_a(p) = 0$ if and only if that curve is supersingular mod p, which happens for an infinite set of p of density zero.

Solved also by the proposer.

Some Inequalities Involving Elementary Symmetric Functions

6629 [1990, 350]. *Proposed by W. E. Briggs, University of Colorado, Boulder, CO.*

Let σ_i be the ith elementary symmetric function of n positive numbers.

(i) Prove that

$$3(\sigma_{i-1}\sigma_{i+2} + \sigma_i\sigma_{i+1})^2 \geqslant 4(\sigma_{i-1}\sigma_{i+1}^3 + \sigma_{i-1}^2\sigma_{i+2}^2 + \sigma_i^3\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$.

(ii)* Is it true that

$$\sigma_{i-1}\sigma_{i+2}^2 + \sigma_i^2\sigma_{i+3} + \sigma_{i+1}^3 > \sigma_{i+1}(\sigma_{i-1}\sigma_{i+3} + 2\sigma_i\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$?

(iii)* Is it true that

$$\sigma_i^2(\sigma_i^2 - \sigma_{i-1}\sigma_{i+1}) > \sigma_{i-1}^2(\sigma_{i+1}^2 - \sigma_i\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$?

Solution of (i) by the proposer. By a theorem of Malo and Schur (see §16 of M. Marden, *Geometry of Polynomials*, Amer. Math. Soc., Providence, 1966 or N. Obreschkoff, *Verteilung und Berechnung der Nullstellen reeller Polynome*, VEBDV, Berlin, 1963) if the zeros of the polynomials $a_0 + a_1x + \cdots + a_nx^n$ and $b_0 + b_1x + \cdots + b_nx^n$ are all real, and if those of one are all of the same sign, then the zeros of $a_0b_0 + a_1b_1x + \cdots + a_nb_nx^n$ are all real. Let $r_1, r_2, \ldots, r_n$ be the n positive numbers of the problem and apply the theorem to

$$(1 + x)^n = 1 + \binom{n}{n-1}x + \cdots + \binom{n}{1}x^{n-1} + x^n$$

and

$$\prod_{k=1}^{n}(x + r_k) = \sigma_n + \sigma_{n-1}x + \cdots + \sigma_1x^{n-1} + x^n;$$

hence the zeros of

$$P(x) = \sigma_n + \binom{n}{n-1}\sigma_{n-1}x + \cdots + \binom{n}{1}\sigma_1x^{n-1} + x^n$$

are all real. Now form $Q(x, y) = y^nP(x/y)$ which is homogeneous of degree n. A partial derivative of $Q(x, y)$ of any order less than or equal to $n - 1$ will have zeros x/y that are all real (by a repeated application of Rolle's theorem). The various partial derivatives of $Q(x, y)$ of order $n - 3$ are, up to a constant factor, of the form $\sigma_{i-1}x^3 + 3\sigma_ix^2y + 3\sigma_{i+1}xy^2 + \sigma_{i+2}y^3$ for some i with $1 \leqslant i \leqslant n - 2$. Since the case $i = n - 1$ (of the proposed problem) is immediate, the result of the problem follows by calculating the discriminant of this cubic polynomial, which must be non-negative. (Note that by adding and subtracting a suitable multiple of $a_0^2a_3^2$ the discriminant of $a_0x^3 + 3a_1x^2 + 3a_2x + a_3$ may be written as

$$27\left[(3a_1^2a_2^2 + 6a_0a_1a_2a_3 + 3a_0^2a_3^2) - 4(a_0a_2^3 + a_0^2a_3^2 + a_1^3a_3)\right].$$

Editorial comment. The inequalities of this problem arose in connection with a paper of the proposer published in the *Rocky Mountain J. of Math.* 15 (1985)

75–89. This paper in turn seems related to "Critical points and sigmoidicity of positive rational functions," by W. G. Bardsley and R. M. W. Wood, this MONTHLY 92 (1985), 37–48.

No other solution of (i) and no solutions at all of (ii) and (iii) were received.

An Assertion of Almost Everywhere Convergence

6630 [1990, 350]. *Proposed by David Pagus, Portland, OR.*

Suppose that $a_1, a_2, a_3, \ldots$ are positive real numbers such that

$$\sum_{n=1}^{\infty} a_n \log(1 + 1/a_n) < \infty.$$

(i) If $\{b_n\}_{n=1}^{\infty}$ is an arbitrary sequence of real numbers, prove that the series

$$\sum_{n=1}^{\infty} a_n |x - b_n|^{-1}$$

converges for almost all real x.

(ii) If k is a fixed positive integer and $\{b_n\}_{n=1}^{\infty}$ is an arbitrary sequence of vectors in $\mathbb{R}^k$, prove that the series

$$\sum_{n=1}^{\infty} a_n \|x - b_n\|^{-k}$$

converges for almost all vectors x in $\mathbb{R}^k$. Here $\|\ \|$ is the usual Euclidean norm on vectors in $\mathbb{R}^k$.

Solution by Kenneth Schilling, University of Michigan–Flint. It is not hard to show that $\lim_{n\to\infty} a_n = 0$, and to deduce from this that

$$\sum_{n=1}^{\infty} a_n < \infty.$$

We establish (ii) directly, since part (i) is simply the case $k = 1$ of (ii). Let $R > 0$ be arbitrary and x a random variable that is uniformly distributed in the ball $B = B_R$ of radius R centered at the origin. Let A denote the volume of the unit ball in $\mathbb{R}^k$, so that B_R has volume AR^k. We shall prove that

$$\sum_{n=1}^{\infty} a_n \|x - b_n\|^{-k} < \infty$$

with probability 1. Since R is arbitrary, this suffices to prove (ii).

We may assume without loss of generality that $|a_n| \leqslant 2R$ for all n, since $|a_n| > 2R$ for at most finitely many n. Less obviously, we may assume $\|b_n\| \leqslant 2R$ for all n. To see this, note that $\|b_n\| > 2R$ implies $\|x - b_n\| \geqslant R$ since $\|x\| \leqslant R$. Hence we have

$$\sum{}' a_n \|x - b_n\|^{-k} \leqslant \sum a_n R^{-k} < \infty$$

for the sum over all n with $\|b_n\| > 2R$.

Now define random variables Y_n by

$$Y_n = \begin{cases} \|x - b_n\|^{-k} & \text{if } \|x - b_n\|^k \geqslant a_n, \\ 0 & \text{otherwise,} \end{cases}$$

and observe that

$$P_n := \text{Prob}\left(Y_n \neq \|x - b_n\|^{-k}\right) \leqslant A\left(a_n^{1/k}\right)^k/(AR^k) = a_n R^{-k}.$$

Since the sum of all P_n is finite, the Borel-Cantelli lemma tells us that

$$\text{Prob}\left(Y_n = \|x - b_n\|^{-k} \text{ for all but finitely many } n\right) = 1.$$

Hence it suffices to show $\Sigma a_n Y_n < \infty$ with probability 1. Now $E(Y_n)$, the expectation of Y_n, is

$$E(Y_n) = \int \|t - b_n\|^{-k}\, dV,$$

where the integral is extended over all t in B_R with $\|t - b_n\|^k \geqslant a_n$. Since all such t clearly satisfy $\|t - b_n\| \leqslant 3R$, the expectation $E(Y_n)$ is bounded above by the above integral extended over all t such that

$$a_n \leqslant \|u\|^k \leqslant (3R)^k,$$

where $u = t - b_n$. By the obvious change of variable,

$$E(Y_n) \leqslant \int \|u\|^{-k}\, dV,$$

where the integral extends over the region between two concentric spheres given by the above u-inequality. Let us change to k-dimensional spherical coordinates; then dV has the form $\rho^{k-1} B(\theta_1, \ldots, \theta_{k-1})\, d\rho\, d\theta_1 \cdots d\theta_{k-1}$, where $B(\theta_1, \ldots, \theta_{k-1})$ is a bounded function of the angular variables $\theta_1, \ldots, \theta_{k-1}$ and $\theta_1, \ldots, \theta_{k-1}$ run over a bounded region. Hence

$$E(Y_n) \leqslant K\left\{\log(3R) + k^{-1} \log(1/a_n)\right\},$$

where K is a constant. Thus

$$E\left(\sum a_n Y_n\right) \leqslant K \log(3R) \sum a_n + Kk^{-1} \sum a_n \log\left(\frac{1}{a_n}\right) < \infty$$

and the result follows.

Editorial comment. Jeffrey Vaaler informs us that similar results are valid in a non-Archimedean setting. For example, in the field of p-adic numbers $\mathbb{Q}_p$ a similar argument yields the following

THEOREM. *Let* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *be sequences in* $\mathbb{Q}_p$ *with each* $a_n \neq 0$. *If*

$$\sum_{n=1}^{\infty} |a_n|_p \log\left(1 + \frac{1}{|a_n|_p}\right) < \infty,$$

then for μ_p-almost all x in $\mathbb{Q}_p$ the infinite series (of real numbers)

$$\sum_{n=1}^{\infty} \frac{|a_n|_p}{|x - b_n|_p}$$

converges. Here μ_p is a Haar measure on $\mathbb{Q}_p$.

Solved also by Jean-Marie Monier (France), José Humberto Ferreira Rosa (Brazil), Jeffrey Vaaler, and the proposer.

Some Unimodal Reciprocal Polynomials with Positive Coefficients

6631 [1990, 432]. *Proposed by Ronald Evans, University of California at San Diego, and Peter Montgomery, Unisys and University of California at Los Angeles.*

We say that $f(z) = a_0 + a_1 z + \cdots + a_k z^k$ is a unimodal reciprocal polynomial with positive coefficients if

$$0 < a_0 = a_k < a_1 = a_{k-1} < a_2 = a_{k-2} < \cdots < a_{\lfloor k/2 \rfloor} = a_{\lfloor (k+1)/2 \rfloor}.$$

(i) If $z_1, z_2, \ldots, z_n$ are the nth roots of 1 and if $2\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients. Here $-\pi < \arg z \leqslant \pi$.

(ii) If $z_1, z_2, \ldots, z_n$ are the nth roots of -1 and if $\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients.

Solution by the proposers. We introduce a real parameter θ that will later be set equal to π/n. For now, however, simply assume that k is a nonnegative integer and that none of $\sin\theta, \sin 2\theta, \ldots, \sin k\theta$ are zero. Define

$$\begin{aligned} P_k(z) &= 1 + \frac{\sin k\theta}{\sin\theta} z + \frac{\sin k\theta \sin(k-1)\theta}{\sin\theta \sin 2\theta} z^2 + \cdots + z^k \\ &= \sum_{j=0}^{k} \frac{\sin k\theta \sin(k-1)\theta \cdots \sin(k+1-j)\theta}{\sin\theta \sin 2\theta \cdots \sin j\theta} z^j. \end{aligned}$$

LEMMA. *If k is odd, then*

$$P_k(z) = (z+1) \prod_{\substack{2 \leqslant j \leqslant k-1 \\ j \text{ even}}} (z^2 + 2z\cos j\theta + 1);$$

if k is even, then

$$P_k(z) = \prod_{\substack{1 \leqslant j \leqslant k-1 \\ j \text{ odd}}} (z^2 + 2z\cos j\theta + 1).$$

then for μ_p-almost all x in $\mathbb{Q}_p$ the infinite series (of real numbers)

$$\sum_{n=1}^{\infty} \frac{|a_n|_p}{|x - b_n|_p}$$

converges. Here μ_p is a Haar measure on $\mathbb{Q}_p$.

Solved also by Jean-Marie Monier (France), José Humberto Ferreira Rosa (Brazil), Jeffrey Vaaler, and the proposer.

Some Unimodal Reciprocal Polynomials with Positive Coefficients

6631 [1990, 432]. *Proposed by Ronald Evans, University of California at San Diego, and Peter Montgomery, Unisys and University of California at Los Angeles.*

We say that $f(z) = a_0 + a_1 z + \cdots + a_k z^k$ is a unimodal reciprocal polynomial with positive coefficients if

$$0 < a_0 = a_k < a_1 = a_{k-1} < a_2 = a_{k-2} < \cdots < a_{\lfloor k/2 \rfloor} = a_{\lfloor (k+1)/2 \rfloor}.$$

(i) If $z_1, z_2, \ldots, z_n$ are the nth roots of 1 and if $2\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients. Here $-\pi < \arg z \leqslant \pi$.

(ii) If $z_1, z_2, \ldots, z_n$ are the nth roots of -1 and if $\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients.

Solution by the proposers. We introduce a real parameter θ that will later be set equal to π/n. For now, however, simply assume that k is a nonnegative integer and that none of $\sin\theta, \sin 2\theta, \ldots, \sin k\theta$ are zero. Define

$$\begin{aligned} P_k(z) &= 1 + \frac{\sin k\theta}{\sin\theta} z + \frac{\sin k\theta \sin(k-1)\theta}{\sin\theta \sin 2\theta} z^2 + \cdots + z^k \\ &= \sum_{j=0}^{k} \frac{\sin k\theta \sin(k-1)\theta \cdots \sin(k+1-j)\theta}{\sin\theta \sin 2\theta \cdots \sin j\theta} z^j. \end{aligned}$$

LEMMA. *If k is odd, then*

$$P_k(z) = (z+1) \prod_{\substack{2 \leqslant j \leqslant k-1 \\ j \text{ even}}} (z^2 + 2z\cos j\theta + 1);$$

if k is even, then

$$P_k(z) = \prod_{\substack{1 \leqslant j \leqslant k-1 \\ j \text{ odd}}} (z^2 + 2z\cos j\theta + 1).$$

Proof. Set $q = e^{2i\theta}$. Define

$$(u)_m = (1-u)(1-qu)(1-q^2u)\cdots(1-q^{m-1}u)$$

for integers $m \geqslant 0$ and complex u. Then

$$P_k(z) = \left(-zq^{(1-k)/2}\right)_k.$$

A famous theorem of Cauchy [**1**, p. 36] states that

$$(u)_k = \sum_{v=0}^{k} \frac{(q)_k}{(q)_v(q)_{k-v}} q^{v(v-1)/2}(-u)^v;$$

(this reduces to the binomial theorem when $q \to 1$). Thus

$$P_k(z) = \sum_{v=0}^{k} \frac{(q)_k}{(q)_v(q)_{k-v}} q^{v(v-k)/2} z^v.$$

The coefficient of z^v in $P_k(z)$ is

$$q^{v(v-k)/2} \prod_{j=1}^{v} \frac{1-q^{k+1-j}}{1-q^j} = \prod_{j=1}^{v} \frac{q^{-(k+1-j)/2} - q^{(k+1-j)/2}}{q^{-j/2} - q^{j/2}}$$

$$= \prod_{j=1}^{v} \frac{\sin(k+1-j)\theta}{\sin j\theta}.$$

Thus the lemma is proved.

We now set $\theta = \pi/n$. To avoid trivial cases we assume $n \geqslant 4$. Let $f(z) = a_0 + a_1 z + \cdots a_k z^k$ be the product considered in the problem. Since monic polynomials with the same roots are identical, $f(z) = P_k(z)$. Since $k < n$, the values $\sin\theta, \ldots, \sin k\theta$ are all positive; in particular, they do not vanish. Hence the Lemma applies and

$$a_v = \prod_{j=1}^{v} \frac{\sin(k+1-j)\theta}{\sin j\theta}.$$

The rest is straightforward. For example,

$$\frac{a_v}{a_{v-1}} = \frac{\sin(k+1-v)\theta}{\sin v\theta} = 1 + 2\frac{\sin((k+1-2v)/2)\theta\cos((k+1)/2)\theta}{\sin v\theta}$$

so $a_{v-1} \leqslant a_v$ if $1 \leqslant v \leqslant [k/2]$. Equality is possible only if $\cos((k+1)/2)\theta = 0$, i.e., $k = n-1$. This is excluded by hypothesis.

Editorial comment. The above proof of the Lemma is due to Evans. Peter Montgomery produced a more elementary proof that uses no more than the addition formulas for sine and cosine, together with induction on k.

If $z_1, \ldots, z_n$ are all the zeros of any monic polynomial of degree n with nonnegative coefficients, it is natural to ask whether the product of the problem

has again nonnegative coefficients. The proposers remark that an affirmative answer is given in [2].

REFERENCES

1. G. Andrews, The Theory of Partitions, Addison-Wesley, 1976.
2. R. Barnard, W. Dayawansa, K. Pearce and D. Weinberg, Polynomials with nonnegative coefficients, to appear in *Proc. Amer. Math. Soc.*

No other solutions were received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Pearls in Graph Theory, a Comprehensive Introduction. By Nora Hartsfield and Gerhard Ringel. Academic Press, San Diego, 1990, ix + 246 pp.

JOAN P. HUTCHINSON
Mathematics and Computer Science Department, Macalester College, St. Paul, MN 55105

Who needs another book on graph theory? I do, if it's a book as refreshingly different as *Pearls in Graph Theory* by Nora Hartsfield and Gerhard Ringel. For example, the opening illustrations of graphs include the San Francisco area road system connecting Half Moon Bay and the Bay Bridge, the chemical graph for aspirin, and the details of a printed circuit board. The book contains a variety of nicely presented topics in graph theory, some classical and some quite unusual. And best of all, it contains a clear and lively three-chapter introduction to the central topics of topological graph theory.

Pearls in Graph Theory begins informally and at an elementary level, suitable for a substantial freshman-sophomore course. After intuitive introductions, concepts and theory are developed with increasing depth, leading into material for a good intermediate-level course. Included also are appropriate open conjectures on, for example, the Oberwohlfach problem, magic and antimagic graphs, and colorings of "earth-moon" maps (more details follow).

In addition, for me the final three chapters are a splendid, enticingly elementary yet comprehensive introduction to topological graph theory, much of which has been pioneered by Ringel. Essentially no other introductory text presents the depth, breadth, and fun of topological graph theory as does this book. (I should confess now that my main love is topological graph theory. What a good way to get student converts to the field, or, more generally, to topology!)

But why and when should one teach any kind of graph theory course and specifically one such as that proposed in this book? I contend that graph theory at the elementary (but also intermediate and advanced) level is an excellent training ground for mathematics: The field is visualizable and concrete; it calls for the development of geometric, algebraic, and logical skills; and even in a one-semester course it can illustrate and point toward both research mathematics and applications modeling.

Consider the case of antimagic graphs: A graph with e edges is called *antimagic* if its edges can be labelled each with a distinct element from $\{1, 2, \ldots, e\}$ so that the sums of the edge labels at each vertex differ. (When the sums are all equal, the graph is called *magic*, a graph theorist's version of a magic square.) This book includes a variety of examples and exercises on antimagic (and magic) labellings, giving ideas to play with, even on the simple structure of trees. The investigations within the text and those suggested as exercises lead to the authors' conjectures that, except for the two-vertex graph with one edge, every tree is antimagic as is every connected graph. Now that looks like a problem at least pieces of which a student should be able to tackle. (Settled cases are given in the exercises.) Magic

and antimagic labellings lead into the important topics of *Leech* and *graceful* labelings of trees.

Thus the field of graph theory, especially as presented in the Hartsfield–Ringel book, provides an ideal training ground for the mathematical process of investigation, generalization, and conjecture, leading to the discovery of proofs or counterexamples. The book contains a pleasant variety of simply stated and intriguing problems, such as the Oberwohlfach problem and the planar coin problem. The former, posed by Ringel after he organized sessions at the Oberwohlfach mathematics institute, asks for seating plans for n meals so that the $2n + 1$ participants have n different dining neighbors. (Solved cases are included.) The coin problem asks: Which planar graphs can be realized with each vertex as a (nonoverlapping) coin in the plane so that adjacent vertices correspond to touching coins? It is conjectured that every planar graph is a coin graph. This work also leads into some recent and intricate work in geometric realizations of graphs, a tantalizing area, but with few proposed techniques for further work.

In summary, I feel that an elementary or intermediate course, for which this book would be appropriate, can serve first to intrigue potential students and then to lead them into a significant body of mathematics, resolved and unresolved. The references include both original research articles and graph theory texts; I would have preferred more pointers toward the intermediate-level references that are included in the reference section. For instance, the Hungarian algorithm to solve the assignment problem is presented, but not proved correct, with references to the papers of Kuhn, Egerváry, and König. The validation of this algorithm is a nice application of Hall's Theorem, as shown in [3].

The placement of a graph theory course, as proposed by this book, may not be completely clear for those of us who now teach (with enthusiasm!) a discrete mathematics course at a similar level. Such a course typically has considerable overlap with the Hartsfield–Ringel graph theory course (e.g., trees, graphs, and related algorithms), but also includes topics such as the proof techniques of induction and contradiction, the theory of sets, subsets, and permutations, and analysis of algorithms. The Hartsfield–Ringel book assumes and uses induction without ado, and in my experience induction is first taught in discrete mathematics courses.

Pearls in Graph Theory has much to offer a solid intermediate-level course, although there are some topics where the instructor may want to introduce and develop more rigor (for example, in the sketch of Prüfer's proof of Cayley's Spanning Tree Formula). The authors do especially nice work with "double counting," a combinatorialist's favorite trick of counting pairs of objects twice, as, for example, when they count the number of paths of k vertices within a complete graph. It is a little disappointing that only two of the four graph algorithms presented come with a proof of correctness. The full proof of Kruskal's algorithm may be intricate for a beginner, but the special case when all edge weights are distinct is easily verified (see [1]).

One of the most significant graph theory achievements in the past 25 years is the proof of the Heawood Map Color Theorem. In 1890 P. J. Heawood, frustrated by inability to prove the Four Color Conjecture, proved that every map on the sphere plus g handles, $g > 0$, can be colored with at most

$$h = \left\lfloor \frac{7 + \sqrt{48g + 1}}{2} \right\rfloor$$

colors, and conjectured that this bound could always be achieved by a map of h mutually adjacent countries. Initially only specific instances were known to be true, and his entire conjecture was finally proved in 1968 by G. Ringel and J. W. T. Youngs in a masterful piece of work (without use of a computer) assisted in parts by Terry, Meyer, and Gustin. This result was a major breakthrough in topological graph theory for the interest it stimulated, for its applicability, and for the development of the Ringel–Youngs' "current graph" technique. Until now, this work and its important techniques have been available primarily to eager graduate students and researchers in [**4**, **5**, **6**], for example, but the ideas involved, when well explained, are straightforward and involve the kind of hands-on experimentation and discovery that comprise much research. I have long wanted to teach the rudiments of topological graph theory and of the Heawood Map Color Theorem, in particular, and *Pearls* provides the means. The last chapter includes a careful and elegant development of the heart of the Ringel–Youngs proof that is a joy to read and would be fun to teach. In fact, the next time I teach topology I think that this work will be my goal. (Such an approach is consistent with the topology textbook by Armstrong [**2**].)

Heawood also made other important topological conjectures on which Ringel and many of his students have worked and generalized. For example, Heawood proposed the following *Empire Problem*. How many colors are needed for a planar map if all countries in the same empire receive the same color (and adjacent countries still receive different colors)? He showed that if each empire contains m countries (and so is dubbed an *m-pire* by Ringel), then the map can be $6m$-colored. B. Jackson and Ringel showed 94 years later that the bound of $6m$ is best possible for all $m > 1$. In a similar vein, Ringel proposed the *Earth-Moon Problem*, which asks for the minimum number of colors needed to color maps on the earth and on a colonized moon so that each country and its lunar colony receive the same color. Such maps can always be 12-colored, but it is unknown whether more than 9 colors are ever needed. On these topics illustrative examples are presented as is introductory theory. My only quibble would be with the proof of the (planar) Five Color Theorem: The authors prove that the countries of every normal (i.e., connected, bridgeless, cubic) planar map can be 5-colored, and it seems to me an easier exercise (building on their work in Chapter 2) to prove that every planar graph can be five-colored. But then again, it was in the context of maps, rather than graphs, that Ringel achieved his celebrated results.

I suppose many of us have contemplated someday writing up our favorite results of graph theory, but how many of us will ever do so? We are most fortunate that as Professor Ringel approaches retirement, he and Nora Hartsfield have made the effort to write up many of their interests in an expository style so that we can share their insights with our present and future students.

REFERENCES

1. M. O. Albertson and J. P. Hutchinson, Discrete Mathematics with Algorithms, Wiley, New York, 1988.
2. M. A. Armstrong, Basic Topology, Springer-Verlag, New York, 1983.
3. J. A. Bondy and U. S. R. Murty, Graph Theory with Applications, American Elsevier, New York, 1976.
4. J. L. Gross and T. W. Tucker, Topological Graph Theory, Wiley, New York, 1987.
5. G. Ringel, Map Color Theorem, Springer-Verlag, New York, 1976.
6. A. T. White, Graphs, Groups and Surfaces (revised edition), North-Holland, Amsterdam, 1984.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P, L. *Nine Papers from the International Congress of Mathematicians 1986.* I.G. Bashmakova, *et al.* Transl. Ser. 2, V. 147. AMS, 1990, vii + 100 pp, $54. [ISBN: 0-8218-3133-X] Translations of papers that appeared in Russian in the *Proceedings* of the 1986 ICM. LAS

General, S. *1-2-3 For Scientists and Engineers, Second Edition.* William J. Orvis. Sybex, 1991, xxiv + 371 pp, $24.95 (P). [ISBN: 0-89588-733-9] Text promotes the idea that spreadsheet software provides an ideal platform for the day-to-day computations and analysis of data in an engineer's life. This edition exploits developments in the Lotus 1-2-3 software since 1987. Little to no theory, but lots of examples and applications. JO

General, P, L.** *Mathematical Impressions.* Anatoliĭ T. Fomenko. AMS, 1990, 184 pp, $45. [ISBN: 0-8218-0162-7] A collection of nearly eighty evocative drawings and oil paintings conveying Fomenko's romantic imaginings ("the freeing of fantasies") in starkly beautiful visual representations. The multi-talented mathematician-artist ("an academic in whose spirit lives a poet") uses topological and analytic surfaces to develop stunning landscapes and intricate images inspired both by mathematical theorems and human emotion. The perfect mathematical coffee-table book. LAS

Mathematics Appreciation, S(12-14). *The Magic of Number.* K.R. Imeson. Mathematical Assoc (Leicester, England), 1989, 96 pp, £3.50 (P). [ISBN: 0-9514716-0-0] A series of brief mathematical "byways," mostly about numbers, derived from talks given over many years to Mathematics Clubs in British schools. Examples: golden proportion, equiangular spiral, irrational numbers, continued fractions, calculation short cuts, primes. Good source for high school projects. LAS

Mathematics Appreciation, T(13-14: 1, 2), L. *Personal Mathematics and Computing: Tools for the Liberal Arts.* Frank Wattenberg. New Liberal Arts Ser. MIT Pr, 1990, xiv + 556 pp, $29.95. [ISBN: 0-262-23157-3] The goal of the course envisioned by the author is to teach students "how to use mathematics to reason about a variety of real, important problems." Using True BASIC as its computing tool, the text investigates problems in probability, economics, optics, distribution of money, and population dynamics. The author leaves students with enough tools and enough open questions after each section to make it plausible that this book might help achieve the author's goals. The problems considered are, for the most part, real and important enough to make this book worth considering for a liberal arts mathematical reasoning course. JO

Precalculus, T(13: 1, 2). *Functions and Graphs, Second Edition.* R. David Gustafson, Peter D. Frisk. Brooks/Cole, 1991, xiii + 559 pp, $49.50. [ISBN: 0-534-14286-9] A simple, straightforward, thorough introduction to standard topics of precalculus mathematics: elementary functions, analytic geometry, polar coordinates, rudiments of complex numbers, matrices and linear equations, Gaussian elimination. Sequences, series, and mathematical induction are briefly surveyed. Elementary functions (algebraic, exponential, logarithmic, and trigonometric) and their graphs are strongly emphasized. Many "story" problems require simple applications of elementary functions and their properties. Other exercises and examples employ calculators or computers. PZ

Finite Mathematics, T(13: 1). *Finite Mathematics.* André L. Yandl. Brooks/Cole, 1991, xxx + 592 pp, $50. [ISBN: 0-534-14190-0] Opens with review and discussion of basic algebra including equations and inequalities in one variable, functions and graphs, then covers standard topics such as matrices, linear programming, probability, and game theory. Many examples worked out in full detail. LC

Education, P. *From the File Treasury. Selections from the Arithmetic Teacher: Mathematics Education through the Middle Grades, 1977-1991.* Ed: Jean M. Shaw. NCTM, 1991. Several dozen K-8 classroom activities reprinted from *The Arithmetic Teacher* on 4x6 cards, arranged in a file by topic: numeration, estimation, fractions, problem solving,

geometry, measurement, time and money, graphing and statistics. Examples: use crackers to explore tessellations of the plane; analyze data from frisbee throws. LAS

History, L. *Operations Analysis in the U.S. Army Eighth Air Force in World War II.* Charles W. McArthur. History of Math., V. 4. AMS, 1990, xxiv + 349 pp, $81. [ISBN: 0-8218-0158-9] A telling of an important untold story—how a group of university mathematicians ("most of whom had never been near a military airplane") helped increase the accuracy of allied bombing from 15% in 1942 to 60% two years later. More a chronology of events and people (many of whom subsequently became very well known) than of mathematical problems, but nonetheless interesting and valuable, especially for insights into the origins of operations research. LAS

Logic, S*(13-18), L. *The Elements of Logic: For Use in Computer Science, Mathematics, and Philosophy.* William J. Edgar. Macmillan, 1989, ix + 150 pp, (P). [ISBN: 0-02-331485-0] Intended for a three or four week component of a course which uses some basic mathematical logic. Written in a slightly informal manner. All standard basic topics included and treated in an elementary manner. Many exercises, all with solutions. Glossary. Suggestions for further reading. Index. RJA

Logic, P. *Lecture Notes in Mathematics-1429: Logic and Computer Science.* S. Homer, *et al.* Springer-Verlag, 1990, 162 pp, $18 (P). [ISBN: 0-387-52734-6] Collection of lecture notes for four short courses given at the C.I.M.E. meeting held June 1988 in Montecatini, Italy. RJA

Logic, T*(17-18: 1, 2), S, P. *Non-Monotonic Reasoning: Formalization of Commonsense Reasoning.* Witold Łukaszewicz. Ser. in AI. Ellis Horwood (US Distr: Prentice Hall), 1990, 328 pp. [ISBN: 0-13-624446-7] Begins with a collection of needed background material including classical propositional, first-order, resolution method, second-order, and modal logic. General discussion of non-monotonic reasoning and its use in artificial intelligence ensues. Bulk of text devoted to modal non-monotonic logics, default logic, circumscription, and the closed world assumption. Bibliography; index. RJA

Logic, S(16-18), P. *Proof Theory and Logical Complexity, Volume I.* Jean-Yves Girard. Stud. in Proof Theory Mono. Bibliopolis (US Distr: Elsevier Science), 1987, 503 pp, $102.50. [ISBN: 0-444-98715-0] Topics include Hilbert's program, semantics of the sequent calculus, applications of the Hauptsatz of Gentzen, Π^1_1 formulas and well-foundedness, the ω-rule, the ordinal ϵ_o, and arithmetic. Begins with background and notation. Note price. Bibliography and index. RJA

Logic, S(14-16), L. *Deducibility and Decidability.* R.R. Rockingham Gill. Routledge, Chapman & Hall, 1990, viii + 162 pp, $47.50. [ISBN: 0-415-00033-5] Intended for students of philosophy. Chapters on the natural numbers, addition, multiplication and exponentiation, theory of Σ-forms, effectiveness, recursively defined predicates, arithmetization of syntax, diagonalization, Gödel's theorems, limits of representability, and arithmetical hierarchy. Appendix; references; index. RJA

Foundations, T(13-14: 1). *Bridge to Abstract Mathematics: Mathematical Proof and Structures, Second Edition.* Ronald P. Morash. McGraw-Hill, 1991, xviii + 389 pp, $42.95. [ISBN: 0-07-043043-8] Changes include reorganization and rewriting of chapters 3 and 4, new emphasis on writing, different design and typeface for "user friendly" appearance, and inclusion of some biographical sketches of famous mathematicians. (*First Edition,* TR, April 1988.) LC

Graph Theory, P. *Graphs and Polyhedra: Binary Spaces and Cutting Planes.* A.M.H. Gerards. CWI Tract, V. 73. Stichting Mathematisch Centrum, 1990, 188 pp, Dfl. 49 (P). [ISBN: 90-6196-390-7] A research monograph applying regular matroid theory, linear programming, and cutting planes to combinatorial optimization problems. Results on signed graphs (respectively, undirected graphs), with no odd-K_4 and no odd-K_3^2 (respectively, and no odd-prism). JPH

Discrete Mathematics, T(13-14: 1, 2). *Discrete Mathematics and Its Applications, Second Edition.* Kenneth H. Rosen. McGraw-Hill, 1991, xxii + 788 pp, $45.95. [ISBN: 0-07-053744-5] Covers logic, sets, methods of proof, algorithms, elementary and advanced counting techniques, relations, graph theory and trees, Boolean algebra, formal languages, and finite-state machines. Numerous exercises and computer projects. Algorithms in both English and pseudocode. More in-depth applications in Michaels and Rosen's book *Applications of Discrete Mathematics.* AD

Discrete Mathematics, P, L*. *Discrete Mathematics across the Curriculum, K-12, 1991 Yearbook.* Margaret J. Kenney, Christian R. Hirsch. NCTM, 1991, viii + 248 pp, $18. [ISBN: 0-87353-305-4] Twenty-eight chapters giving a wide variety of teaching strategies and topical examples of how to introduce discrete mathematics—"the math for our time"—into the K-12 school curriculum as called for in the NCTM *Standards.* A rich source of ideas for teachers who are beginning to develop curricula in this area. LAS

Linear Algebra, T(14: 1). *Elementary Linear Algebra with Applications, Second Edition.* Richard O. Hill, Jr. Harcourt Brace Jovanovich, 1991, xv + 492 pp, $33 net. [ISBN: 0-15-521092-0] This edition offers an optional early treatment of determinants; adds about 200 exercises, including some cumulative review exercises; adds five new sections, including applications of determinants and iterative methods; and makes some editorial changes. (1986 Academic Press *First Edition,* TR, October 1986.) OJ

Group Theory, S(18), P. *Automorphic Representations of Unitary Groups in Three Variables.* Jonathan D. Rogawski. Annals of Math. Stud., V.

123. Princeton Univ Pr, 1990, xii + 259 pp, $19.95 (P); $60. [ISBN: 0-691-08587-0; 0-691-08586-2] Automorphic representations of the group $G = U(3)$ with basic field F are studied by means of comparison of trace formulas of G and Res G for a quadratic extension E of F. This involves a stabilization procedure initially due to Langlands and developed here with applications. References, indexes. JS

Algebra, P. *First Siberian Winter School: Algebra and Analysis.* Eds: A.D. Aleksandrov, *et al.* Transl. Ser. 2, V. 148. AMS, 1991, vii + 112 pp, $63. [ISBN: 0-8218-3700-1] Ten advanced expository papers based on lectures delivered in March 1987 at Kemerovo State University intended to acquaint analysts with the current frontiers of algebra. LAS

Algebra, P. *Semigroup Algebras.* Jan Okniński. Pure & Appl. Math., V. 138. Marcel Dekker, 1990, ix + 357 pp, $119.50. [ISBN: 0-8247-8356-5] A unified presentation of recent results in the theory of noncommutative semigroup rings over a field. Receiving special attention are semigroup algebras of cancellative semigroups, semigroup algebras satisfying polynomial identities, and finiteness conditions on semigroup algebras. SG

Algebra, S(18), P. *Ideals of Identities of Associative Algebras.* Aleksandr Robertovich Kemer. Transl. of Math. Mono., V. 87. AMS, 1991, v + 81 pp, $75. [ISBN: 0-8218-4548-9] After an introduction summarizing developments in the theory of algebras defined by polynomial identities, the author leads into his own recent work, presenting several major results. Most notable is a solution of Specht's problem with a proof that any associative algebra over a field of characteristic zero has a finite basis of identities. Bibliography, index. JS

Algebra, T(15-16: 2), L. *Abstract Algebra.* Dennis Kletzing. Harcourt Brace Jovanovich, 1991, xii + 605 pp, $35 net. [ISBN: 0-15-500391-7] Groups, rings, fields including sections on p-groups and Sylow theory, the geometric groups, Galois theory and solvability by radicals. Copious examples, appendices on history of groups, rings and fields, glossary. LC

Differential Equations, T(18: 1), P. *Lecture Notes in Mathematics-1445: Regularity Theory for Quasilinear Elliptic Systems and Monge—Ampère Equations in Two Dimensions.* Friedmar Schulz. Springer-Verlag, 1990, xv + 123 pp, $16 (P). [ISBN: 0-387-53103-3] Self-contained introduction to characteristic theory for the named equations. Presents the Heinz-Lewy theory plus auxiliary material. Also includes basic results about Hölder continuity and regularity theory for linear elliptic equations. BL

Differential Equations, P. *Exponential Sums and Differential Equations.* Nicholas M. Katz. Annals of Math. Stud., No. 124. Princeton Univ Pr, 1990, xi + 430 pp, $22.50 (P); $65. [ISBN: 0-691-08599-4; 0-691-08598-6] Relates the theory of linear differential equations in one complex variable with polynomial coefficients to the theory of one-parameter families of exponential sums over finite fields. JO

Differential Equations, P. *Exterior Differential Systems.* R.L. Bryant, *et al.* Math. Sci. Res. Inst., V. 18. Springer-Verlag, 1991, vii + 475 pp, $49. [ISBN: 0-387-97411-3] Comprehensive reference on the subject, with applications to partial differential equations and linear differential operators. This book is not an introduction. JO

Partial Differential Equations. *Surveys on Analysis, Geometry and Mathematical Physics.* Eds: Bert-Wolfgang Schulze, Hans Triebel. Teubner-Texte zur Math., Band 117. BG Teubner Leipzig, 1990, 308 pp, (P). [ISBN: 3-322-00773-1] Five papers on boundary value problems, non-linear problems, and pseudo-differential operators. LAS

Partial Differential Equations, P. *Solitons in Molecular Systems, Second Edition.* A.S. Davydov. Transl: Eugene S. Kryachko. Math & Its Applic., V. 61. Kluwer Academic, 1991, xiii + 413 pp, $149. [ISBN: 0-7923-1029-2] Presents the theory of excited states in quasi-one-dimensional periodic structures using the theory of non-linear differential equations. Background theory in first five chapters includes discussion of role of solitons in biological systems such as muscle contraction and general anesthesia. JO

Partial Differential Equations, P. *Multidimensional Hyperbolic Problems and Computations.* Eds: James Glimm, Andrew J. Majda. IMA, V. 29. Springer-Verlag, 1991, xiv + 386 pp, $39.80. [ISBN: 0-387-97485-7] Proceedings of an IMA workshop held April 3-14, 1989 as part of the program on nonlinear waves. Twenty-six papers covering a wide cross-section of research: kinetic theory of gases, systems of hyperbolic conservation laws, nonlinèar geometric optics, shock waves, and numerical methods. BL

Partial Differential Equations, P. *Bounded and Almost Periodic Solutions and Nonlinear Operator Differential Equations.* A.A. Pankov. Math. & Its Applic., V. 55. Kluwer Academic, 1990, x + 220 pp, $99. [ISBN: 0-7923-0585-X] Given a bounded function from a locally compact group to a Banach space and a positive number epsilon, an element of the group is an epsilon-almost period of the function if the supremum taken over the group of the difference between the function and its translate by that element is less than epsilon. A function is almost periodic if each subset of epsilon-almost periods is relatively dense in the group. Treats solutions of evolution equations and of spacially almost periodic problems. BL

Partial Differential Equations, P. *Hamiltonian Flows and Evolution Semigroups.* Henryk Gzyl. Pitman Res. Notes in Math. Ser., V. 239. Longman Sci & Tech (US Distr: Wiley), 1990, 124 pp, $32 (P). [ISBN: 0-582-03190-7] Develops a non-unitary representation theory for canonical transformations of Hamiltonian systems and an operational calculus similar to that in quantum mechanics. The representation theory maps semigroups and special functions associated with different Hamiltonians onto one another. The operational calculus is used to asso-

ciate an evolution semigroup to each Hamiltonian. A variety of examples illustrate the results and methods. BL

Partial Differential Equations, P. *Lagrangian Manifolds and the Maslov Operator.* A.S. Mishchenko, V.E. Shatalov, B. Yu. Sternin. Transl: Dana Mackenzie. Springer-Verlag, 1990, x + 395 pp, $69.50. [ISBN: 0-387-13613-4] Presents a method for obtaining global asymptotic solutions to partial and pseudodifferential equations of physical interest, e.g., quantum mechanics, wave propagation. Using geometric optics, the amplitude of a wave may become infinite where rays focus or form caustics. The Maslov operator translates the existence and uniqueness of solutions to conditions on cocycles on a submanifold in phase space. Part I develops the necessary topological tools. Part II constructs the canonical operator and presents applications. BL

Partial Differential Equations, P. *Inverse Source Problems.* Victor Isakov. Math. Surveys & Monographs, No. 34. AMS, 1990, xiv + 193 pp, $72. [ISBN: 0-8218-1532-6] Treats inverse problems for all three major types of partial differential equations, focusing on elliptic equations. Contains a basic review of the direct problem and an overview of inverse problems, uniqueness, existence, and regularity theorems. Also treats stability and numerical solutions. BL

Numerical Analysis, P. *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations.* Eds: Roland Glowinski, *et al.* SIAM, 1991, xi + 417 pp, $56.50 (P). [ISBN: 0-89871-278-5] Proceedings of a symposium held in Moscow, May 21-25, 1990. Thirty-three papers on the method of "solving boundary value problems which allow concurrent computations (parallel processing) to occur on different subdomains." BL

Numerical Analysis, P, L. *Finite Element Methods (Part 1).* Eds: P.G. Ciarlet, J.L. Lions. Handbook of Numer. Analysis, V. II. North-Holland (US Distr: Elsevier Science), 1991, ix + 928 pp, $110. [ISBN: 0-444-70365-9] Contains six articles entitled "Finite Elements: An Introduction," "Basic Error Estimates for Elliptic Problems," "Local Behavior in Finite Element Methods," "Mixed and Hybrid Methods," "Eigenvalue Problems," and "Evolution Problems." AO

Numerical Analysis, P. *Advances in Numerical Partial Differential Equations and Optimization.* Eds: S. Gómez, J.P. Hennart, R.A. Tapia. SIAM, 1991, xii + 365 pp, $47.50 (P). [ISBN: 0-89871-269-6] Proceedings of the Fifth Mexico-United States Workshop on Numerical Analysis in Merida, Yucatan, Mexico, January 1989. Numerical aspects of optimization, linear algebra, and differential equations. MLR

Functional Analysis, P. *Littlewood-Paley Theory and the Study of Function Spaces.* Michael Frazier, Björn Jawerth, Guido Weiss. CBMS Reg. Conf. Ser. in Math., No. 79. AMS, 1991, vii + 132 pp, $40 (P). [ISBN: 0-8218-0731-5] Fifteen lectures from a 1989 conference at Auburn presenting the basics of Littlewood-Paley theory, which provides a unifying framework for a variety of seemingly unrelated function spaces. JO

Analysis, P. *Mathematical Problems of Tomography.* Eds: I.M. Gelfand, S.G. Gindikin. Transl. of Math. Mono., V. 81. AMS, 1990, iii + 267 pp, $96. [ISBN: 0-8218-4534-9] Translations from Russian of eight papers related to computerized tomography, the mathematics of CT-scans, and of many applications in radio astronomy and electron micrography. The papers are further unified by their attention to the mathematical restrictions arising from the practical nature of the problems. JO

Algebraic Geometry, P. *Algebraic Geometry: Sundance 1988.* Eds: Brian Harbourne, Robert Speiser. Contemp. Math., V. 116. AMS, 1991, xi + 146 pp, $57 (P). [ISBN: 0-8218-5124-1] A collection of ten papers on algebraic curves and related varieties. SG

Algebraic Geometry, P. *Fewnomials.* A.G. Khovanskiĭ. Transl. of Math. Mono., V. 88. AMS, 1991, viii + 139 pp, $96. [ISBN: 0-8218-4547-0] Concerned with the general problem of estimating the number of positive solutions of a system of k real polynomials in k variables in terms of k and the number of monomials that appear in the polynomials. Complex and transcendental versions are also presented. SG

Differential Geometry, T?(17: 2). *Basic Elements of Differential Geometry and Topology.* S.P. Novikov, A.T. Fomenko. Math. & Its Applic., V. 60. Kluwer Academic, 1990, zx + 490 pp, $164. [ISBN: 0-7923-1009-8] Based on a series of courses at Moscow State University, this book is intended to present those aspects of geometry and topology that are essential for a modern mathematical education. Includes classical theory of curves and surfaces, tensor analysis, and basic homotopy. Seven appendices provide the preceding theory with some life in the form of examples. Note the price. JO

Differential Geometry, T(17-18: 1), S. *Manifold Theory: An Introduction for Mathematical Physicists.* Daniel Martin. Math. & Its Applic. Ellis Horwood, 1991, 423 pp, $81. [ISBN: 0-13-543877-2] Differential geometry text for graduate students in physics. Standard topics. Very readable—could be used for independent study. Many examples in the first chapters, only a few from physics. A few proofs in the later chapters are omitted. Answers, hints, or partial solutions to about half the exercises. Bibliography. OJ

Geometry. *Mathographics.* Robert Dixon. Dover, 1991, ix + 214 pp, $8.95 (P). [ISBN: 0-486-26639-7] Over 100 mathematical drawings created by compass constructions or computer graphics. Intended to intrigue and inspire the reader to replicate or extend the drawings. Some figures are explained; many are not. Minimal narrative provides mathematical background, but the heart of the book is in the artwork. MW

Algebraic Topology, T(18: 1), S, P. *Topology of 4-Manifolds.* Michael H. Freedman, Frank Quinn. Math. Ser., V. 39. Princeton Univ Pr, 1990, viii + 259 pp, $49.50. [ISBN: 0-691-08577-3] Masterful presentation of the basic tools and some major results in the study of 4-manifolds. Offers simplified treatments of the embedding theorems for 2-disks, the classification theorem for closed 1-connected 4 manifolds, and more general classification theorems. Almost a third of the material is new. Most chapters conclude with references for the major ideas and results they contain. A strong algebraic (geometric) topology course is a prerequisite. OJ

Operations Research, P, L. *Mathematical Developments Arising from Linear Programming.* Eds: Jeffrey C. Lagarias, Michael J. Todd. Contemp. Math., V. 114. AMS, 1990, xiv + 341 pp, $59 (P). [ISBN: 0-8218-5121-7] Proceedings of the AMS Summer Research Conference held at Bowdoin College in 1988. Contains twenty-one papers dealing mostly with various aspects of interior-point linear programming algorithms. Nonlinear optimization and integer programming problems are also discussed. These papers represent current research and are neither introductions to nor surveys of linear programming. SM

Operations Research, T(16-17), L. *Game Theory: Analysis of Conflict.* Roger B. Myerson. Harvard Univ Pr, 1991, xiii + 568 pp. [ISBN: 0-674-34115-5] A very well-written introduction to game theory presented in the style of a mathematics text with many theorems and a heavy use of mathematical symbols. The ideas, techniques, and notation from undergraduate courses in calculus, linear algebra, and probability are assumed throughout the text. Exercises appear after each of the ten chapters; none of the exercises are solved. Extensive bibliography. SM

Optimization, T(16-17: 1), C, P, L.** *Knapsack Problems: Algorithms and Computer Implementations.* Silvano Martello, Paolo Toth. Wiley, 1990, xii + 296 pp, $110. [ISBN: 0-471-92420-2] Presents state-of-the-art exact and approximate algorithms for a number of important *NP*-hard integer programming problems: binary, bounded, unbounded, and binary multiple knapsack problems as well as subset-sum, change-making, generalized assignment, and bin-packing problems. Includes theoretical and empirical results on computational complexity. Fortran codes for the most effective algorithms are provided on a diskette included with the book. AO

Systems Theory, P. *Progress in Fuzzy Sets and Systems.* Eds: Wolfgang H. Janko, Marc Roubens, H.-J. Zimmermann. Theory & Decision Lib., Ser. D, V. 5. Kluwer Academic, 1990, ix + 188 pp, $87. [ISBN: 0-7923-0730-5] Contains the proceedings of a workshop held April 6-8, 1989 in Vienna. Fifteen refereed papers, in alphabetical order. RJA

Stochastic Processes, P. *Lecture Notes in Mathematics-1464: Ecole d'Eté de Probabilités de Saint-Flour XIX–1989.* D.L. Burkholder, E. Pardoux, A. Sznitman. Springer-Verlag, 1991, vi + 259 pp, $30 (P). [ISBN: 0-387-53841-0] Three sets of course lectures: D. Burkholder on martingale theory; E. Pardoux on *filtrage stochastiques* (in French); and A. Sznitman on propagation of chaos. (One signature, pp. 58-90, bound up-side-down.) LAS

Computational Statistics, P. *The Frontiers of Expert Systems and Artificial Intelligence.* Ed: Edward J. Dudewicz. Amer. Ser. in Math. & Management Sci., V. 27. American Sciences Pr, 1989, 214 pp, $98.75 (P). [ISBN: 0-935950-29-X] Contains nine papers presented at the First International Conference on Statistical Computing held March 30-April 2, 1987 in Çeşme, Izmir, Turkey. RJA

Statistics, P. *COMPSTAT 1990: Proceedings in Computational Statistics.* Eds: K. Momirović, V. Mildner. Springer-Verlag, 1990, xiii + 336 pp, $69 (P). [ISBN: 0-387-91390-4] A collection of forty-nine papers from the ninth COMPSTAT Symposium held at Dubrovnik, Yugoslavia, 1990. Topics include multivariate data analysis, statistical expert systems, and robust computing. RWJ

Statistics, P. *Statistical Multiple Integration.* Eds: Nancy Flournoy, Robert K. Tsutakawa. Contemp. Math., V. 115. AMS, 1991, xii + 276 pp, $71 (P). [ISBN: 0-8218-5122-5] Papers from a 1989 conference of mathematicians, statisticians, and computational scientists on problems and techniques in statistical multiple integration, primarily arising from multivariate and Bayesian statistics. RM

Programming, T(15-17: 1), L*. *An Introduction to Object-Oriented Programming.* Timothy Budd. Addison-Wesley, 1991, xv + 399 pp, $36.75. [ISBN: 0-201-54709-0] A language-independent introduction to the concepts of object-oriented programming illustrated with examples from C++, Smalltalk, Objective-C, and Object Pascal. Case studies present practical examples of the use of object-oriented programming techniques. Covers advanced topics such as metaclasses, delegation, parameterized classes, and multiple inheritance as well as implementation of object-oriented languages. AO

Algorithms, P*, L.** *Handbook of Algorithms and Data Structures In Pascal and C, Second Edition.* G.H. Gonnet, R. Baeza-Yates. Intern. Comput. Sci. Ser. Addison-Wesley, 1991, xiv + 424 pp, $34.50. [ISBN: 0-201-41607-7] This new edition (*First Edition,* TR, February 1985) incorporates many recent results and a new chapter on text searching. The bibliography contains references to thirty-six textbooks and 1350 published research papers! Gives Pascal and/or C implementations of many of the algorithms discussed in the text. AO

Algorithms, T(17-18: 1, 2), P, L. *Analysis of Algorithms and Data Structures.* Lech Banachowski, Antoni Kreczmar, Wojciech Rytter. Intern. Comput. Sci. Ser. Addison-Wesley, 1991, xii + 300 pp, $36.75. [ISBN: 0-201-41693-X] Presents a variety of approaches to the analysis of algorithms and data structures including probabilistic models, generating functions, embeddings between data structures, the

pointer machine model of computations, and pebble games for modeling space-efficient computations. Also discusses transformational techniques for speeding up algorithms (dynamic and algebraic simulation) and analysis of parallel algorithms. AO

Computer Systems, S?(13), L. *Life with UNIX: A Guide For Everyone.* Don Libes, Sandy Ressler. Prentice Hall, 1989, xx + 346 pp. [ISBN: 0-13-536657-7] Chatty guide (not a tutorial or manual!) to UNIX: how to use it, history, folk features that are rarely documented, all the neat stuff that "everybody knows" but that no one who is new can find out except by asking a guru. RM

Computer Systems, P. *UNIX Curses Explained.* Berny Goodheart. Prentice Hall, 1991, xiii + 287 pp, (P). [ISBN: 0-13-931957-3] The UNIX *curses* library is a set of routines that facilitate the writing of interactive screen-based programs. This volume provides a tutorial on the use of the *curses* library, and can also function as a reference manual for the library as implemented in UNIX System V.3. AO

Computer Systems, P. *C Library Reference for UNIX System V.4 (with ANSI C).* Specialized Systems Consultants (POB 55549, Seattle, WA 98155), 1991, 72 pp, $8 (P). [ISBN: 0-916151-47-6] A pocket reference guide to the UNIX System V, Release 4 version of the C programmer's library. ANSI prototypes are used to indicate the calling conventions of the functions, and functions that are part of the ANSI standard library are flagged. AO

Computer Graphics, P. *Scientific Visualization and Graphics Simulation.* Ed: Daniel Thalmann. Wiley, 1990, ix + 264 pp, $49.95. [ISBN: 0-471-92742-2] Eighteen lectures originally given at the Swiss Federal Institute of Technology, each written by one or several experts in a particular aspect of scientific visualization. The lectures present techniques of computer graphics, image synthesis, and animation as well as applications in engineering, science, and medicine. AO

Theory of Computation, S(17-18), P. *Logic and Computation: Interactive Proof with Cambridge LCF.* Lawrence C. Paulson. Tracts in Theoretical Comput. Sci., V. 2. Cambridge Univ Pr, 1990, xiii + 302 pp, $24.95 (P). [ISBN: 0-521-39560-7] LCF (Logic for Computable Functions) denotes computer programs for reasoning about computation. The design of Cambridge LCF (based on Edinburgh LCF) is explained here. Part 1 contains background in elementary logic and domain theory. Part 2 explains the features and details of the program. Bibliography and index. (*First Edition,* TR, June-July 1989.) RJA

Artificial Intelligence, T(15-17: 1, 2), S, L. *Knowledge and Inference.* Makoto Nagao. Transl: Richard Weyrauch, Yasuko Kitajima. Academic Pr, 1990, xv + 304 pp, $42.50. [ISBN: 0-12-513662-5] An attempt to merge artificial intelligence and library science is presented. Discusses the current state of library science and the place of artificial intelligence in it. Includes search and problem solving, methods of making proofs, and the use of knowledge in looking for a proof. Finishes with a chapter on expert systems. Chapter summaries, key word lists, exercises. Answers to exercises. List of recommended books by topics. Index. RJA

Computer Science, S(17-18), P. *Implementing Persistent Prolog: Large, Dynamic, Shared Procedures in Prolog.* Robert M. Colomb. Ser. in Comput. & Applic. Ellis Horwood (US Distr: Prentice Hall), 1990, 199 pp. [ISBN: 0-13-453531-6] Begins with a simple interpreter and the principle issues in implementing Prolog. The next chapters include persistence research, handling large procedures, persistent dynamic procedures, shared procedures, types, the Warren Abstract Machine, NU-Prolog, and integration of a persistent procedure management system onto existing implementations. Extensive lists of figures and of tables. References, index. RJA

Applications, P. *Proceedings of the Second Workshop on Road-Vehicle-Systems and Related Mathematics.* Ed: Helmut Neunzert. ECMI, V. 4. Kluwer Academic, 1989, 235 pp, $79. [ISBN: 0-7923-0243-5] Twelve papers dealing with various aspects of road-vehicle systems. Mathematical descriptions of roads, vehicles, vehicle dynamics, road reactions, and reliability are presented. These papers would be most beneficial to researchers in the area and their advanced students. SM

Applications (Fluid Dynamics), T(18: 2), L. *Computational Techniques for Fluid Dynamics 1: Fundamental and General Techniques, Second Edition.* C.A.J. Fletcher. Ser. in Computat. Physics. Springer-Verlag, 1991, xiii + 401 pp, $49 (P). [ISBN: 0-387-53058-4] Intended to introduce engineering graduate students to techniques of computational fluid dynamics. Includes brief differential equations review, discussion of discretization techniques, and presentation of general techniques for diffusion and convection equations. *Volume 2* applies these general techniques to specific problems. IBM disk with Fortran programs available from the author. JO

Applications (Physics), P. *Quantum Mechanics, Algebras and Distributions.* D.A. Dubin, M.A. Hennings. Pitman Res. Notes in Math. Ser., V. 238. Longman Sci & Tech (US Distr: Wiley), 1990, 238 pp, $38 (P). [ISBN: 0-582-06776-6] Presents a rigorous formulation of quantum mechanics based on the algebra of observables and states. Intended to provide rigorous background for students and researchers who might otherwise do calculational research without mastery of the deeper theory. JO

Reviewers

RJA: Richard J. Allen, St. Olaf; LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; JPH: Joan P. Hutchinson, Macalester; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; BL: Brian Loe, Carleton; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; JS: John Schue, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

Foundation Chair of Mathematics Education

Applications are invited for appointment to the Foundation Chair of Mathematics Education within the Institute of Education.

The University seeks a leader in the field of mathematics education, with an outstanding research record and wide recognition in the field, strong qualifications in mathematics, and an excellent record as a teacher.

The appointee will build up a centre for those wishing to develop expertise or undertake further research in mathematics education.

The chair will be located within the School of Science and Mathematics Education in the Institute of Education. The new professor will strengthen co-operative links already established with the Department of Mathematics in the Faculty of Science.

This is a continuing chair, with an initial salary of AUD$72,000 rising to AUD$76,000 in July 1992.

Further information about the position, including details of application procedure, conditions for outside work, superannuation, travel and removal expenses, housing assistance and conditions of appointment, is available from the Registrar. All correspondence (marked "PERSONAL AND CONFIDENTIAL") should be addressed to the Registrar, The University of Melbourne, Parkville, Victoria, 3052, Australia.

Telephone enquiries: (613) 344 7527. Facsimile: (613) 344 6897.

Applications close on 31 January, 1992.

The Council reserves the right to make no appointment or to fill the position by invitation at any stage.

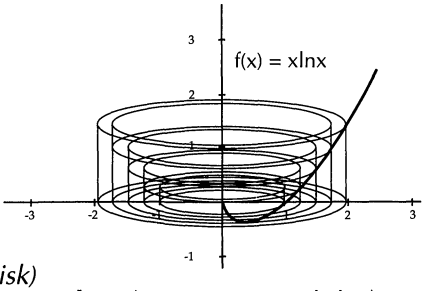
The University of Melbourne is an equal opportunity employer and has implemented a smoke-free workplace policy.

THE UNIVERSITY OF MELBOURNE

arm51375

SHOULD CALCULUS & PRECALCULUS BE TAUGHT DIFFERENTLY?

Software **POWERFUL DISCOVERY TOOL**



BestGrapher Software allows students to take control of their own learning through a hands-on approach. **BestGrapher** is easy to learn and easier to use.

- Function plotting and finding function zeros
- Symbolic differentiation and numerical integration
- Draws secants, tangents and normals
- Approximates volumes of solids of revolution
- Graphs polar equations and parametric equations
- *Available for IBM or Macintosh (Specify 3.5 or 5.25 disk)*
- *Quantity pricing, LAN licenses are available*

BestGrapher: $73.00 (Demo disk: $10.00)

Book **COMPUTER LABORATORY GUIDE FOR CALCULUS**

Using BestGrapher is a rich selection of exercises and laboratory investigations designed to supplement any calculus textbook, and any calculus software. The chapters contain a variety of challenging, non-routine computer exercises.

- Problems that encourage experimentation and discovery about functions, derivatives, areas and other topics in calculus and precalculus
- Lab projects that contain open-ended questions, promote working in small groups, and culminate in a written report.
- Work-text format with answers to selected problems

All prices include shipping & handling **Using BestGrapher**: $18.00 (Spiral bound, 150 pages)

Venture Publishing 9 Bartlet St., Suite 55 Andover, MA 01810 508-475-3733

Winning Women into Mathematics

Patricia Clark Kenschaft, Editor

American media often ask why women "can't" do mathematics. Any answer is misleading. Better questions are needed, along with indications of how to find potential answers.

The Committee on the Participation of Women of the Mathematical Association of America was established in 1987 "to work for full involvement of women in MAA activities that will encourage women to pursue careers in the mathematical sciences." With this book, the Committee seeks to expand the number and effectiveness of those winning women into mathematics. WINNING WOMEN is written to inform, to empower, and to inspire.

The Committee identifies fifty-five cultural customs that discourage aspiring women mathematicians. They tell us how these customs can be changed and what can be done to recruit, retain, and acknowledge women in mathematics. A bibliography of over 100 sources on the issues of women's participation in mathematics is included, as well as descriptions of programs that have been successful in encouraging young women to study mathematics. The book is filled with interesting anecdotes, and contains over 50 photographs of prominent women in mathematics.

88 pp., 1991 , Paperbound
ISBN 0-88385-453-8

List: $11.00 MAA Member: $9.00

Catalog Number: WIW

CONTENTS

- A bibliography of over 100 sources on the issues of women's participation in mathematics
- Fifty-five cultural patterns causing American women to be underrepresented in mathematics
- What you personally can do
- Programs that succeed
- A history of women in mathematics-especially in the MAA
- A chronicle of the programs, articles, and suggestions of the Committee on the Participation of Women
- A minority woman's viewpoint
- An overview of the statistics
- Photographs, anecdotes, cartoons

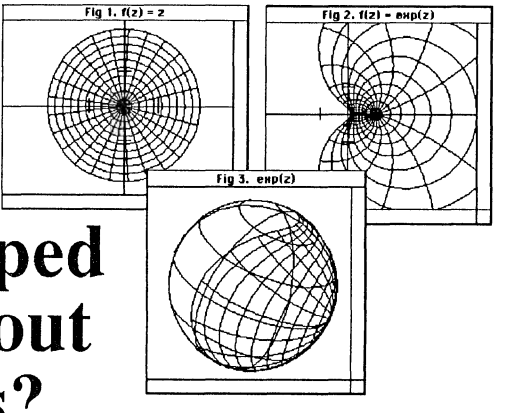


ORDER FROM:

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036

(FAX) (202) 265-2384
Prepaid orders sent postage & handling free. Visa and Mastercard orders accepted. (Please give the card number and expiration date on credit card orders) We will bill for orders over $10.00.

When was the last time a computer program helped you think about mathematics?



Software and video tapes for the student, professional, and anyone who loves mathematics. ***Ask for our latest catalog.***

Lascaux Graphics (602) 544-4229 **(800) 338-0993**
7601 N. Calle Sin Envidia, Suite 31 - Tucson, AZ 85718 USA

THE MATHEMATICAL SYMBOL FOR QUALITY

New from McGraw-Hill for 1992

CALCULUS AND ANALYTIC GEOMETRY
Fifth Edition
Stein/Barcellos

DISCOVERING CALCULUS WITH THE HP 28 AND HP 48
Smith/Minton, Softcover

INTRODUCTION TO MATHCAD™ FOR SCIENTISTS AND ENGINEERS
(also available with the Student Edition of MathCAD 2.5)
Wieder, Softcover

CALCULUS GEMS
Simmons, Softcover

FOUNDATIONS OF ABSTRACT MATHEMATICS
Kurtz

*MathCAD™ is a trademark of MathSoft, Inc.

CALCULUS FOR BUSINESS, ECONOMICS, AND THE SOCIAL AND LIFE SCIENCES
Fifth Edition
Hoffmann/Bradley

ENCOUNTERS WITH CHAOS
Gulick

NUMERICAL METHODS AND ANALYSIS
Buchanan/Turner

STATISTICAL METHODS IN THE BIOLOGICAL AND HEALTH SCIENCES
Second Edition
Milton

FUNDAMENTALS OF COMPUTING: VOLUME I LOGIC, PROGRAMS, AND COMPUTING
Tucker/Bradley/Cupper/Garnick

Also Available

APPLICATIONS OF DISCRETE MATHEMATICS
Michaels/Rosen

DISCRETE MATHEMATICS AND ITS APPLICATIONS, 2/E
Rosen

DIFFERENTIAL EQUATIONS WITH APPLICATIONS AND HISTORICAL NOTES, 2/e
Simmons

PARTIAL DIFFERENTIAL EQUATIONS AND BOUNDARY VALUE PROBLEMS WITH APPLICATIONS
Pinsky

For more information please contact your McGraw-Hill representative or write to McGraw-Hill College Division, Comp Processing and Control, P.O. Box 448, Hightstown, NJ 08520-0448.

ZENITH DATA SYSTEMS PRESENTS
MASTERS OF INNOVATION

Solving The Problem Of How Students Solve Problems.



Dr. Ronald H. Stevens

Any professor will tell you that good teaching requires an understanding of their students' thinking. At the University of California Los Angeles, however, Dr. Ronald H. Stevens has found an especially innovative way to get inside the cranium of his second-year medical students.

With his award-winning "IMMEX" software program, Dr. Stevens discovers not only *if* his students can solve Immunology problems, but also *how* information was gathered and processed during the solving.

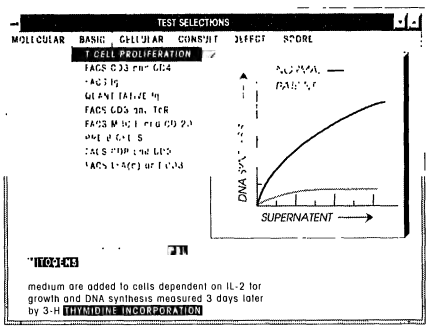
Here's how it works. Programmed in Microsoft® *Windows™ version 3.0,* the easy-to-use "IMMEX" consists of multiple cases of immune defects and a set of results from 45 laboratory tests (see figure below). Through a series of exercises and exams, students are asked to diagnose these cases by selecting the appropriate tests and examining their results.

Upon completion, graphical representations are generated by computer to show which tests were chosen and importantly, demonstrate how students searched for the solution to each problem.

It is then possible to visualize the students' thought process in a way that standard, multiple choice testing doesn't allow. For instance, Dr. Stevens can learn how organized and focused their knowledge is, how well their organization relates to critical concepts in Immunology, where major misconceptions exist and whether proper knowledge links are evident.

In turn, these insights gained through "IMMEX" can have a significant impact on teaching methods. As Dr. Stevens explains, "This approach can lead to rapid detection and remediation of individual students' problem solving difficulties, and can greatly personalize the education process."

What was Dr. Stevens' approach in developing his award-winning software? He chose to program "IMMEX" on Zenith Data Systems laptop PCs. They provided him with all the Random Access Memory and portability required to work after hours and over weekends. And that made solving the problem of how students solve problems, less of a problem.



"IMMEX" Software Program

ABOUT THE MASTERS OF INNOVATION COMPETITION.

As a corporation committed to education, Zenith Data Systems encourages students and educators—like Dr. Ronald H. Stevens —to creatively explore the potential of computers within their fields of study. Towards that end, Zenith Data Systems has sponsored the MASTERS OF INNOVATION Competition for the past three years.

To obtain an unabridged copy of his discussion paper on "IMMEX," or an application to enter the MASTERS OF INNOVATION IV Competition, please write to us at: Masters Of Innovation Program, Zenith Data Systems Corporation, P.O. Box 14513, Chicago, IL 60614-9998.

Microsoft and *Windows* are trademarks of Microsoft Corporation. Copyright © 1991 Dr. Ronald H. Stevens.
Copyright ©1991 Zenith Data Systems Corporation.

Introducing E.Z. Math, E.Z. Algebra and E.Z. Arithmetic for the HP 48SX

E.Z. Math, E.Z. Algebra and E.Z. Arithmetic are programs for the Hewlett Packard 48SX calculator conceived, written and programmed by Raymond La Barbera and the E.Z. Software Company. Each program comes on a 128K plug-in ROM card accompanied by an easy-to-understand, well-written, detailed manual loaded with lots of specific examples and is designed for use by students, teachers. parents and business people. Each program features an easy-to-use, logically organized, user-friendly interface which enables those who consider themselves to be calculator and computer illiterates, as well as those who don't like to read manuals, to have full access to all program features quickly and easily. Since the HP 48SX is essentially an impressive, sharp-looking, 8 ounce pocket computer, students are easily motivated to take it along with them to study, practice, drill and master math in a study hall, on a train or bus, in a car, on line, on vacation, on a break—in short, for self-study at any time and in any place.

What Can Be Done With E.Z. Math

E.Z. Math effectively solves problems involving graphs, numbers, loans and savings. With E.Z. Math, anyone can:

- Master the entire high school and college graphing curriculum, from algebra to calculus, with 188 families of equations, inequalities, functions, and systems, all arranged in an easy-to-use, user-friendly system of menus to make graphic analysis a snap!
- Get extensive help with calculations involving fractions, whole numbers, complex numbers and number sequences.
- Easily do savings and loan calculations and generate complete amortization tables.
- Learn many basic concepts including those involving sets, variables, graphing, solving, numbers, loans and savings.

What Can Be Done With E.Z. Algebra

E.Z. Algebra is a comprehensive ninth grade high school basic algebra course as well as a high school and college remedial algebra course that builds a solid algebra foundation. With E.Z. Algebra, anyone can:

- Learn about sets, operations, variables, relations and other concepts essential to a real understanding of algebra.
- Understand the sets of natural numbers, whole numbers, integers, rational numbers and real numbers.
- Master the meaning and properties of the operations of addition, subtraction, multiplication, division, power and root.
- Do all kinds of problems involving algebraic expressions, numerical phrases, equations and inequalities.

What Can Be Done With E.Z. Arithmetic

E.Z. Arithmetic is a comprehensive elementary school basic arithmetic course as well as a high school and college remedial arithmetic course that makes solving most arithmetic problems a snap! With E.Z. Arithmetic, anyone can:

- Learn how to add, subtract, multiply, divide and order whole numbers, fractions, decimals, percents and integers.
- Master the meaning, terminology and conversion methods for whole numbers, fractions, decimals and percents.
- Drill and be graded on endlessly varied, randomly selected sets of problems involving whole numbers, fractions, decimals and integers, with the difficulty level, number of problems, operation and type of number user selectable.

How To Order Copies or Get Further Information

Each E.Z. Software program costs $130.00 ($125.00 retail, plus $5.00 shipping and handling). Take a 10% discount when ordering ten or more units. We accept payment by check, money order, COD, VISA, MC, AE and purchase order. If within 30 days you find that any E.Z. Software program fails to meet your expectations, we'll gladly take back your copy for a prompt, courteous refund. To order copies, either individually or bundled with HP 48SX calculators, please contact:

SMI Corporation, 250 West New Street, Dept MM2, Kingsport, Tennessee 37660
(800) 234-0123 or (615) 378-4821 or (615) 245-8982 (Fax).

Cambridge University Press

Logic and Information
K. Devlin
1991 336 pp. 41030-4 Hardcover about $34.50

Exact Constants in Approximation Theory
N. Korneichuk
K. Ivanov, *Translator*
Encyclopedia of Mathematics and its Applications 38
1991 464 pp. 38234-3 Hardcover $89.50

Representations of Finite Groups of Lie Type
Francois Digne *and* ***Jean Michel***
London Mathematical Society Student Texts 21
1991 166 pp. 40117-8 Hardcover $44.50
40648-X Paper $17.95

Now in paperback...

A Course in Mathematics for Students of Physics
Volumes 1 and 2
Paul Bamberg *and* ***Shlomo Sternberg***
Vol 1: 424 pp. 40649-8 Paper $27.95
Vol 2: 461 pp. 40650-1 Paper $27.95

Statistical Inference for Spatial Processes
B. D. Ripley
160 pp. 42420-8 Paper $19.95

Twistor Geometry and Field Theory
R. S. Ward *and* ***R. D. Wells, Jr.***
Cambridge Monographs on Mathematical Physics
532 pp. 42268-X Paper $34.50

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.
MasterCard/VISA accepted. Prices subject to change.

NEW! VERSION 2

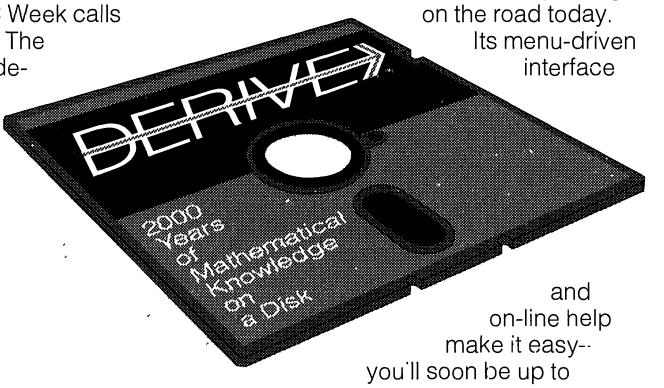
Recursive & iterative programming • Over 70 new math functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE May 29, 1990 Derive, Version 1 6

DERIVE®– A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it *"Editors' Choice."* PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics.



High performance.

DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays ormulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even handheld computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. Its menu-driven interface and on-line help make it easy-- you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

No sticker shock.

DERIVE's suggested retail price is $250. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. Pacific Standard Time.

And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

New from MIT

Foundations of Computing Series, Research Reports and Notes

Basic Category Theory for Computer Scientists
Benjamin C. Pierce
128 pp $17.95 softcover

Categories, Types, and Structures
An Introduction to Category Theory for the Working Computer Scientist
Andrea Asperti and Giuseppe Longo
325 pp $32.50

Logic Programming and Nonmonotonic Reasoning
Proceedings of the First International Workshop
June 22-24, 1991 Washington, D.C.
edited by Anil Nerode, Wiktor Marek, and V. S. Subrahmanian
432 pp $32.50 softcover

Recommended for text use

Linear Network Optimization
Algorithms and Codes
Dimitri P. Bertsekas
This book explains important recent algorithms such as auction and relaxation, proposed by the author and others for the solution of these problems.
275 pp $39.95

The Art of the Metaobject Protocol
Gregor Kiczales, Jim des Rivières, and Daniel Bobrow
335 pp $45.00 cloth, $24.95 softcover

To order call toll-free 1-800-356-0343 Fax orders (617) 625-6660
MasterCard and VISA accepted

The MIT Press
55 Hayward Street, Cambridge, MA 02142

FREE CATALOG!

TO HELP TEACH MATHEMATICS EFFECTIVELY



Request a free copy of the NCTM catalog today! It lists over 180 titles with annotations.

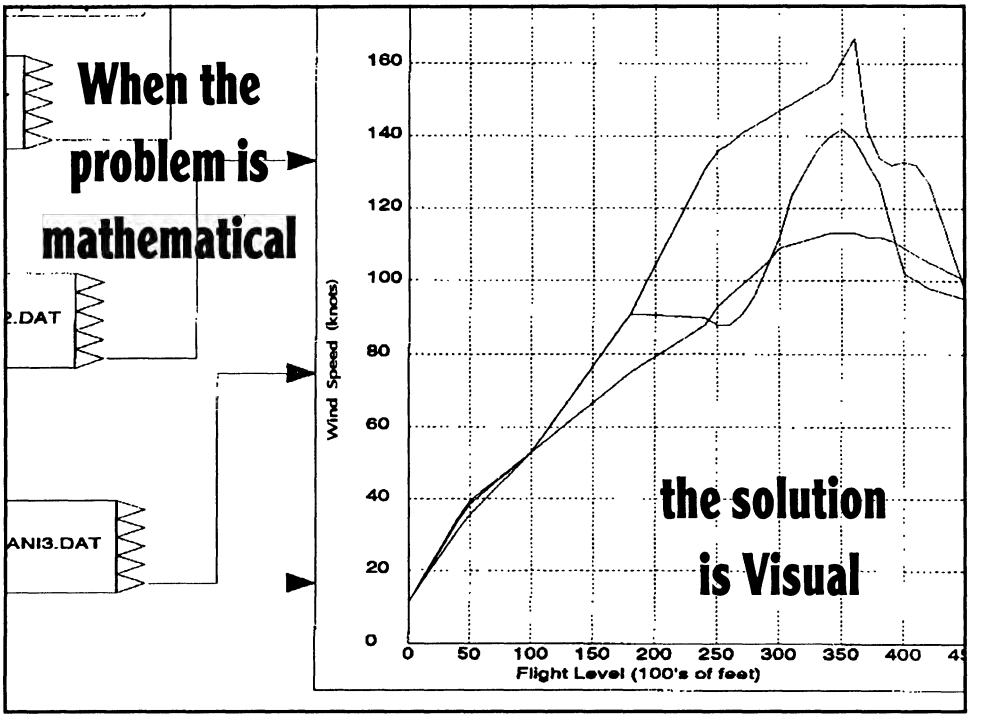
NCTM educational materials have been written by mathematics teachers for mathematics teachers. Each publication focuses on how mathematics can be taught more effectively. Materials are available for all grade levels and are helpful to practicing teachers and future teachers.

Refer these books to your students. They will find many classroom-tested ideas and helpful ways to improve the teaching and learning of mathematics.

If you would like multiple copies of the catalog, let us know.

National Council of Teachers of Mathematics
1906 Association Drive, Reston, VA 22091, Attn: Dept. P-MAA
Tel.(703) 620–9840/ext. 128; fax (703) 476–2970

NCTM



VisSim is the powerful mathematical modeling software tool for the 1990s.

VisSim introduces an unrivaled visual programming technique for high performance scientific analysis.

VisSim numerically solves non-linear differential equations for a multitude of engineering and scientific disciplines.

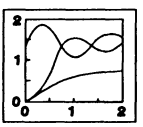
Visual Solutions, Inc.
487 Groton Road
Westford, MA 01886
To order, call (508) 692-5499

VisSim sells for thousands less than its nearest competitor and yet has 65 block functions and a choice of 6 integration methods.

VisSim/RT add-on does data acquisition and control in real time.

VisSim plot options include: *XY, Time History, Log, Log-Log, Autoscaling, Color Coding, Over Plotting, Grid Lines, Labeling, Geometric Markers.*

VisSim is the mathematical modeling software available under Windows 3.0 for only $895. Demos $25.



Visual Solutions

VisSim is a trademark of Visual Solutions Inc.
Windows is a registered trademark of Microsoft Corporation

Two Books On Mathematical Ingenuity By Ross Honsberger

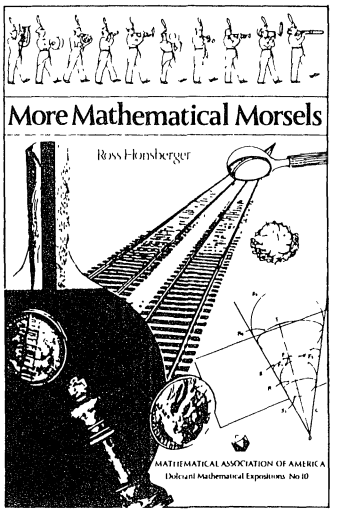
Ross Honsberger is the author of seven books in the Dolciani Mathematical Exposition series, each of which presents problems from algebra, arithmetic, number theory, probability, and geometry, and provides ingenious solutions and/or intriguing results. His most recent addition to the series is MORE MATHEMATICAL MORSELS which is a continuation of the earlier Mathematical Morsels volume published in 1979. Now is your opportunity to own both of these excellent books. The problems presented are meant to be enjoyed, rather than instruct, although instruction is almost always the automatic by-product.

Honsberger says in the Preface of his first volume of morsels, "Mathematics abounds in bright ideas. No matter how long and hard one pursues her, mathematics never seems to run out of exciting surprises. And by no means are these gems to be found only in difficult work at an advanced level. All kinds of simple notions are full of ingenuity. The present volume discusses scores of elementary problems and contains dozens of marvellous ideas."

All of the problems are accessible to anyone with a knowledge of freshman mathematics. Buy both of these excellent books today, and share them with your students.

This is an excellent text...It exposes gifted undergraduate students to a set of elementary problems...whose proofs are ingenious, usually containing beautiful ideas...Its lucid expository style and stimulating mathematical content make its reading a rewarding experience for mathematicians at all levels. (about Mathematical Morsels)

M.S. Cheema in *Mathematical Reviews*



...even the more advanced theoretician will find many of the problems fascinating...

D.T. Swift-Hook in *Physics Education (about Mathematical Morsels)*

NEW NEW NEW NEW NEW NEW

MORE MATHEMATICAL MORSELS

Ross Honsberger

315 pp., Paperbound, 1990,
ISBN 0-88385-313-2

List: $16.00 MAA Member: $13.50

Catalog Number DOL-10

AN MAA BESTSELLER

MATHEMATICAL MORSELS

Ross Honsberger

249 pp., Hardbound, 1979
ISBN 0-88385-303-5

List: $ 28.00 MAA Member: $ 21.00

Catalog Number DOL-03

Both volumes purchased now

List: $40.00 MAA Member: $29.00

Catalog Number DOL-310

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

INDEX TO VOLUME 98, 1991
THE AMERICAN MATHEMATICAL MONTHLY

Title Index	989	Reviews by Author	996
Author Index	991	Index of Special Topics	996
Subject Index	994	Solutions	997
Reviews by Title	995	Problems Proposed	997
		Problems Solved	998

TITLE INDEX

Algebraic Characterization of Compact Abelian Groups, M. V. Deshpande and S. R. Ghorpade, 235

An Analytic Technique to prove Borel's Strong Law of Large Numbers, Liu Wen, 146

An Analytical Description of Some Simple Cases of Chaotic Behaviour, James V. Whittaker, 489

Another Elementary Approach to the Jordan Form, J. I. Hall, 336

Anthropomorphic Polygons, Godfried Toussaint, 31

Arcs with Positive Measure and a Space-Filling Curve, Timothy Lance and E. Thomas, 124

Borromean Circles are Impossible, Bernt Lindström and H.-O. Zetterström, 340

Box-Spline Tilings, Carl de Boor and K. Höllig, 793

Branching Processes That Grow Faster Than Binary Splitting, F. M. Dekking, 728

Carlyle Circles and the Lemoine Simplicity of Polygon Constructions, Duane W. DeTemple, 97

The Cauchy Problem for the Wave Equation with Distribution Data: an Elementary Approach, Calvin H. Wilcox, 401

A Characterization of Baire Spaces, Dragan Janković, M. Ganster, and I. Reilly, 624

A Characterization of the Cantor Function, Donald R. Chalice, 255

Circumscribed Circles, Robert Osserman, 419

A Coloring Problem, Ira M. Gessel, 530

A Continued Fraction and Permutations with Fixed Points, Henri Darmon and J. McKay, 25

Continuous Nowhere-Differentiable Functions—an Application of Contraction Mappings, Hidefumi Katsuura, 417

Convergence-Preserving Functions: An Alternative Discussion, Arthur Smith, 831

The Converse of Liouville's Theorem, Richard Grant, 354

Conway's Challenge Sequence, Colin L. Mallows, 5

Crossing the Square, Keith Whittington, 833

Derangements, Permanents, and Christmas Presents, Stephen G. Penrice, 617

The Derivative á la Carathéodory, Stephen Kuhn, 40

Differentiation of Approximately Specified Functions, C. W. Groetsch, 847

Disks and Shells Revisited, Walter Carlip, 154

Distance Functions and Topologies, Fred Galvin and S. D. Shore, 620

A Dynamic Puzzle, Moshe Rosenfeld, 22

An Early Introduction to Dynamics, Marysia T. Weiss, 635

A Very Elementary proof of the Malgrange–Ehrenpreis Theorem, Jean-Pierre Rosay, 518

On the Equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$, Alfonso Villani, 826

The Exploding Exponential and Other Chaotic Bursts in Complex Dynamics, Robert L. Devaney and M. B. Durkin, 217

On Finite Groups and Finite Fields, J. D. Reid, 549

First-Class Functions, G. Myerson, 237

A Freshman Calculus Proof That Real Quartics Factor, John Isbell, 749

The Generation of All Rational Orthogonal Matrices, Hans Liebeck and A. Osborne, 131

Geographical Boundary Extrema, HelenJane Armstrong and R. Herz-Fischler, 752

The Gibbs Phenomenon for Piecewise-Linear Approximation, J. Foster and F. B. Richards, 47

Gram–Schmidt Orthogonalization by Gauss Elimination, Lyle Pursell and S. Y. Trimble, 544

Hex Games and Twist Maps on the Annulus, Steve Alpern and A. Beck, 803

How Not to Repeatedly Differentiate a Reciprocal, Robert A. Leslie, 732

How Things Float, E. N. Gilbert, 201

How to Pick Out the Integers in the Rationals: An Application of Number Theory to Logic, Dan Flath and S. Wagon, 812

Hypocycloids, Continued Fractions, and Distribution Modulo One, Norman Richert, 133

The Inclusion of $L^p(\mu) \subseteq L^q(\nu)$, A. G. Miamee, 342

Inequalities for Areas Associated with Conics, W. A. Day, 36

On Infinite and Finite Covering Systems, M. A. Berger, A. Feltzenbaum, A. S. Fraenkel, and R. Holzman, 739

Irrationality Without Number Theory, Richard Beigel, 332

L'Hôpital's Rule Via Integration, Donald Hartig, 156

Lecturers on Minimal Surfaces, Johanne C. C. Nitsche, 72

Linear Algebra and Quantum Chemistry, Jerome A. Goldstein and M. Levy, 710

Linear Algebra by Analogy, Scott H. Hochwald, 918

Lines of Best Fit by Graphics and the Wald Line, R. F. Duffin, 835

Magic Squares and Linear Algebra, Christopher J. Henrich, 481

On the Mean Value Theorem, Inequality, and Inclusion, M. Furi and M. Martelli, 840

A Moment Convergence Theorem, Lajos Takács, 742

Monthly Unsolved Problems, 1969–1991, Richard K. Guy, 973

More on Cross Products, Paul Binding, 850

Multiple Pythagorean Number Triples, Albert Fässler, 518

Norm Inequalities for the Powers of a Matrix, Man Kam Kwong and A. Zettl, 533

A Note on Dixon's Proof of Cauchy's Integral Theorem, Peter A. Loeb, 242

A Note on Finding a Strict Saddlepoint, Daniel Bienstock, F. Chung, M. Fredman, A. A. Schäffer, P. W. Shor, and S. Suri, 418

Number Systems with a Complex Base: a Fractal Tool for Teaching Topology, Daniel Goffinet, 249

Number Theory as Gadfly, B. Mazur, 593

Numerical Linear Algebra on the HP-28 or How to Lie with Supercalculators, Yves Nievergelt, 539

An Obvious Proof of Burnside's Lemma, Kenneth B. Bogart, 927

An Occurrence of the Ballot Numbers in Operator Theory, Lawrence H. Riddle, 613

An "Ocean of Zeros" Proof That a Certain Non-Linear Number is Transcendental, M. J. Knight, 947

An Overview of the Calculus Curriculum Reform Effort: Issues for Learning, Teaching, and Curriculum Development, Joan Ferrini-Mundy and K. G. Graham, 627

Parametric Differentiation Revisited, L. R. Bragg, 259

Part metric and Hyperbolic Metric, H. S. Bear, 109

The Perimeter of a Rose, James P. Butler, 139

Poisson Integrals of Riemann Integrable Functions, J. Král and W. F. Pfeffer, 929

Positive Definite Matrices and Sylvester's Criterion, George T. Gilbert, 44

On the Product of Two Alternating Matrices, Dragomir Ž. Doković, 935

Proof of the Recursive Unsolvability of Hilbert's Tenth Problem, James P. Jones and Y. V. Matijasevič, 689

The Ranks of Tournament Matrices, D. de Caen, 829

Reliability, Recursion, and Risk, Lavon B. Page and J. E. Perry, 937

Some Remarks on Weak Continuity, James Chew and J. Tong, 931

On the Right-Hand Derivative of a Certain Integral Function, John Klippert, 751

A New Scheme for Multiple-Choice Tests in Lower-Division Mathematics, Bruce R. Johnson, 427

The Search for a Finite Projective Plane of Order 10, C. W. H. Lam, 305

The Self-Paced Calculus Program at Berkeley, Murray H. Protter, 245

On Series of Partial Fractions, P. D. Barry and D. J. Hurley, 240

A Simple Interpretation of the Complex Contour Integral, Alan Gluchoff, 641

A Simple Proof of the Weierstrass Approximation Theorem, J. H. Lindsey, II, 429

A Simple Proof of Zorn's Lemma, Jonathan Lewin, 353

A Simple Proof of a Simple Version of the Riemann Mapping Theorem by Simple Functional Analysis, P. R. Garabedian, 824

Simple Tests for Classifying Critical Points of Quadratics with Linear Constraints, Paul Binding, 949

Smoothness from Finite Points, Victor Neves, 27

Statistics Within Departments of Mathematics at Liberal Arts Colleges, Thomas L. Moore and J. A. Witmer, 431

Stirling's Series and Bernoulli Numbers, Elias Y. Deeba and D. M. Rodriguez, 423

The Structure of Orthogonal Transformations, Lee Rudolph, 349

Symmetry Classes: Functions of Three Variables, N. Metropolis and G.-C. Rota, 328

Teaching and Testing Mathematics Reading, Carl C. Cowen, 50

$3Q$; $1A$, Henry S. Warren, Jr., 611

A Topological Mean Value Theorem for the Plane, Ira Rosenholtz, 149

A Tree Among the Stepping Stones, A. R. Meijer, 144

Two Equal Sums of Three Squares with Equal Products, John B. Kelly, 527

Unions and Common Complements of Subspaces, K. P. S. Bhaskara Rao and A. Ramachandra Rao, 127

Visualizing the p-adic Integers, Albert A. Cuoco, 355

Which Latin Squares are Cayley Tables?, Man-Keung Siu, 625

Why the Method of Undetermined Coefficients Works, Clay C. Ross, Jr., 747

From Wythoff's Nim to Chebyshev's Inequality, Kenneth B. Stolarsky, 889

On the Zeros of $\tan^{(k)} z$, Li-Chien Shen, 524

Nested Radicals, Jonathan A. Borwein and G. Debarra, 735

π and the Limit of $(\sin \alpha)/\alpha$, Leonard Gillman, 346

A Topological Approach to Digital Topology, T. Yung Kong, R. Kopperman, and P. R. Meyer, 901

AUTHOR INDEX

Alpern, Steve and A. Beck, Hex Games and Twist Maps on the Annulus, 803

Armstrong, HelenJane and R. Herz-Fischler, Geographical Boundary Extrema, 752

Barry, P. D. and D. J. Hurley, On Series of Partial Fractions, 240

Bear, H. S., Part metric and Hyperbolic Metric, 109

Beck, Anatole and S. Alpern, Hex Games and Twist Maps on the Annulus, 803

Beigel, Richard, Irrationality Without Number Theory, 332

Berger, M. A., A. Feltzenbaum, A. S. Fraenkel, and R. Holzman, On Infinite and Finite Covering Systems, 739

Bhaskara Rao, K. P. S. and A. Ramachandra Rao, Unions and Common Complements of Subspaces, 127

Bienstock, Daniel, F. Chung, M. Fredman, A. A. Schäffer, P. W. Shor, and S. Suri, A Note on Finding a Strict Saddlepoint, 418

Binding, Paul, More on Cross Products, 850

Binding, Paul, Simple Tests for Classifying Critical Points of Quadratics with Linear Constraints, 949

Bogart, Kenneth B., An Obvious Proof of Burnside's Lemma, 927

Borwein, Jonathan A. and G. Debarra, Nested Radicals, 735

Bragg, L. R., Parametric Differentiation Revisited, 259

Butler, James P., The Perimeter of a Rose, 139

Carlip, Walter, Disks and Shells Revisited, 154

Chalice, Donald R., A Characterization of the Cantor Function, 255

Chew, James and J. Tong, Some Remarks on Weak Continuity, 931

Chung, Fan, D. Bienstock, M. Fredman, A. A. Schäffer, P. W. Shor, and S. Suri, A Note on Finding a Strict Saddlepoint, 418

Cowen, Carl C., Teaching and Testing Mathematics Reading, 50

Cuoco, Albert A., Visualizing the p-adic Integers, 355

Darmon, Henri and J. McKay, A Continued Fraction and Permutations with Fixed Points, 25

Day, W. A., Inequalities for Areas Associated with Conics, 36

DeTemple, Duane W., Carlyle Circles and the Lemoine Simplicity of Polygon Constructions, 97

Debarra, G. and J. A. Borwein, Nested Radicals, 735

de Boor, Carl and K. Höllig, Box-Spline Tilings, 793

de Caen, D., The Ranks of Tournament Matrices, 829

Deeba, Elias Y. and D. M. Rodriguez, Stirling's Series and Bernoulli Numbers, 423

Dekking, F. M., Branching Processes That Grow Faster Than Binary Splitting, 728

Deshpande, M. V. and S. R. Ghorpade, Algebraic Characterization of Compact Abelian Groups, 235

Devaney, Robert L. and M. B. Durkin, The Exploding Exponential and Other Chaotic Bursts in Complex Dynamics, 217

Doković, Dragomir Ž., On the Product of Two Alternating Matrices, 935

Duffin, R. J., Lines of Best Fit by Graphics and the Wald Line, 835

Durkin, Marilyn B. and R. L. Devaney, The Exploding Exponential and Other Chaotic Bursts in Complex Dynamics, 217

Fässler, Albert, Multiple Pythagorean Number Triples, 518

Feltzenbaum, A., M. A. Berger, A. S. Fraenkel, and R. Holzman, On Infinite and Finite Covering Systems, 739

Ferrini-Mundy, Joan and K. G. Graham, An Overview of the Calculus Curriculum Reform Effort: Issues for

Learning, Teaching, and Curriculum Development, 627

Flath, Dan and S. Wagon, How to Pick Out the Integers in the Rationals: An Application of Number Theory to Logic, 812

Foster, J. and F. B. Richards, The Gibbs Phenomenon for Piecewise-Linear Approximation, 47

Fraenkel, A. S., M. A. Berger, A. Feltzenbaum, and R. Holzman, On Infinite and Finite Covering Systems, 739

Fredman, Michael, D. Bienstock, F. Chung, A. Schäffer, P. W. Shor, and S. Suri, A Note on Finding a Strict Saddlepoint, 418

Furi, M. and M. Martelli, On the Mean Value Theorem, Inequality, and Inclusion, 840

Galvin, Fred and S. D. Shore, Distance Functions and Topologies, 620

Ganster, Maxmillian, D. Janković, and I. Reilly, A Characterization of Baire Spaces, 624

Garabedian, P. R., A Simple Proof of a Simple Version of the Riemann Mapping Theorem by Simple Functional Analysis, 824

Gessel, Ira M., A Coloring Problem, 530

Ghorpade, Sudhir R. and M. V. Deshpande, Algebraic Characterization of Compact Abelian Groups, 235

Gilbert, E. N., How Things Float, 201

Gilbert, George T., Positive Definite Matrices and Sylvester's Criterion, 44

Gillman, Leonard, π and the Limit of $(\sin\alpha)/\alpha$, 346

Gluchoff, Alan, A Simple Interpretation of the Complex Contour Integral, 641

Goffinet, Daniel, Number Systems with a Complex Base: a Fractal Tool for Teaching Topology, 249

Goldstein, Jerome A. and M. Levy, Linear Algebra and Quantum Chemistry, 710

Graham, Karen Geuther and J. Ferrini-Mundy, An Overview of the Calculus Curriculum Reform Effort: Issues for Learning, Teaching, and Curriculum Development, 627

Grant, Richard, The Converse of Liouville's Theorem, 354

Groetsch, C. W., Differentiation of Approximately Specified Functions, 847

Guy, Richard K., Monthly Unsolved Problems, 1969–1991, 973

Höllig, Klaus and C. de Boor, Box-Spline Tilings, 793

Hall, J. I., Another Elementary Approach to the Jordan Form, 336

Hartig, Donald, L'Hôpital's Rule Via Integration, 156

Henrich, Christopher J., Magic Squares and Linear Algebra, 481

Herz-Fischler, Roger and HelenJane Armstrong, Geographical Boundary Extrema, 752

Hochwald, Scott H., Linear Algebra by Analogy, 918

Holzman, R., M. A. Berger, A. Feltzenbaum, and A. S. Fraenkel, On Infinite and Finite Covering Systems, 739

Hurley, D. J. and P. D. Barry, On Series of Partial Fractions, 240

Isbell, John, A Freshman Calculus Proof That Real Quartics Factor, 749

Janković, Dragan, M. Ganster, and I. Reilly, A Characterization of Baire Spaces, 624

Johnson, Bruce R., A New Scheme for Multiple-Choice Tests in Lower-Division Mathematics, 427

Jones, James P. and Y. V. Matijasevič, Proof of the Recursive Unsolvability of Hilbert's Tenth Problem, 689

Katsuura, Hidefumi, Continuous Nowhere-Differentiable Functions—an Application of Contraction Mappings, 417

Kelly, John B., Two Equal Sums of Three Squares with Equal Products, 527

Klippert, John, On the Right-Hand Derivative of a Certain Integral Function, 751

Knight, M. J., An "Ocean of Zeros" Proof That a Certain Non-Linear Number is Transcendental, 947

Kong, T. Yung, R. Kopperman, and P. R. Meyer, A Topological Approach to Digital Topology, 901

Kopperman, Ralph, T. Yung Kong, and P. R. Meyer, A Topological Approach to Digital Topology, 901

Král, J. and W. F. Pfeffer, Poisson Integrals of Riemann Integrable Functions, 929

Kuhn, Stephen, The Derivative á la Carathéodory, 40

Kwong, Man Kam and A. Zettl, Norm Inequalities for the Powers of a Matrix, 533

Lam, C. W. H., The Search for a Finite Projective Plane of Order 10, 305

Lance, Timothy and E. Thomas, Arcs with Positive Measure and a Space-Filling Curve, 124

Leslie, Robert A., How Not to Repeatedly Differentiate a Reciprocal, 732

Levy, Mel and J. A. Goldstein, Linear Algebra and Quantum Chemistry, 710

Lewin, Jonathan, A Simple Proof of Zorn's Lemma, 353

Liebeck, Hans and A. Osborne, The Generation of All Rational Orthogonal Matrices, 131

Lindsey, J. H. II, A Simple Proof of the Weierstrass Approximation Theorem, 429

Lindström, Bernt and H.-O. Zetterström, Borromean Circles are Impossible, 340

Loeb, Peter A., A Note on Dixon's Proof of Cauchy's Integral Theorem, 242

Mallows, Colin L., Conway's Challenge Sequence, 5

Martelli, M. and M. Furi, On the Mean Value Theorem, Inequality, and Inclusion, 840

Matijasevič, Yuri V. and J. P. Jones, Proof of the Recursive Unsolvability of Hilbert's Tenth Problem, 689

Mazur, B., Number Theory as Gadfly, 593

McKay, John and H. Darmon, A Continued Fraction and Permutations with Fixed Points, 25

Meijer, A. R., A Tree Among the Stepping Stones, 144

Metropolis, N. and G.-C. Rota, Symmetry Classes: Functions of Three Variables, 328

Meyer, P. R., T. Yung Kong, and R. Kopperman, A Topological Approach to Digital Topology, 901

Miamee, A. G., The Inclusion of $L^p(\mu) \subseteq L^q(\nu)$, 342

Moore, Thomas L. and J. A. Witmer, Statistics Within Departments of Mathematics at Liberal Arts Colleges, 431

Myerson, G., First-Class Functions, 237

Neves, Victor, Smoothness from Finite Points, 27

Nievergelt, Yves, Numerical Linear Algebra on the HP-28 or How to Lie with Supercalculators, 539

Nitsche, Johanne C. C., Lecturers on Minimal Surfaces, 72

Osborne, Anthony and H. Liebeck, The Generation of All Rational Orthogonal Matrices, 131

Osserman, Robert, Circumscribed Circles, 419

Page, Lavon B. and J. E. Perry, Reliability, Recursion, and Risk, 937

Penrice, Stephen G., Derangements, Permanents, and Christmas Presents, 617

Perry, Jo Ellen and L. B. Page, Reliability, Recursion, and Risk, 937

Pfeffer, W. F. and J. Král, Poisson Integrals of Riemann Integrable Functions, 929

Protter, Murray H., The Self-Paced Calculus Program at Berkeley, 245

Pursell, Lyle and S. Y. Trimble, Gram–Schmidt Orthogonalization by Gauss Elimination, 544

Ramachandra Rao, A. and K. P. S. Bhaskara Rao, Unions and Common Complements of Subspaces, 127

Reid, J. D., On Finite Groups and Finite Fields, 549

Reilly, Ivan, D. Janković, and M. Ganster, A Characterization of Baire Spaces, 624

Richards, F. B. and J. Foster, The Gibbs Phenomenon for Piecewise-Linear Approximation, 47

Richert, Norman, Hypocycloids, Continued Fractions, and Distribution Modulo One, 133

Riddle, Lawrence H., An Occurrence of the Ballot Numbers in Operator Theory, 613

Rodriguez, Dennis M. and E. Y. Deeba, Stirling's Series and Bernoulli Numbers, 423

Rosay, Jean-Pierre, A Very Elementary Proof of the Malgrange–Ehrenpreis Theorem, 518

Rosenfeld, Moshe, A Dynamic Puzzle, 22

Rosenholtz, Ira, A Topological Mean Value Theorem for the Plane, 149

Ross, Clay C., Jr., Why the Method of Undetermined Coefficients Works, 747

Rota, Gian-Carlo and N. Metropolis, Symmetry Classes: Functions of Three Variables, 328

Rudolph, Lee, The Structure of Orthogonal Transformations, 349

Schäffer, Alejandro A., D. Bienstock, F. Chung, M. Fredman, P. W. Shor, and S. Suri, A Note on Finding a Strict Saddlepoint, 418

Shen, Li-Chien, On the Zeros of $\tan^{(k)} z$, 524

Shor, Peter W., D. Bienstock, F. Chung, M. Fredman, A. A. Schäffer, and S. Suri, A Note on Finding a Strict Saddlepoint, 418

Shore, S. D. and F. Galvin, Distance Functions and Topologies, 620

Siu, Man-Keung, Which Latin Squares are Cayley Tables? 625

Smith, Arthur, Convergence-Preserving Functions: An Alternative Discussion, 831

Stolarsky, Kenneth B., From Wythoff's Nim to Chebyshev's Inequality, 889

Suri, Subhash, D. Bienstock, F. Chung, M. Fredman, A. A. Schäffer, and P. W. Shor, A Note on Finding a Strict Saddlepoint, 418

Takács, Lajos, A Moment Convergence Theorem, 742

Thomas, Edward and T. Lance, Arcs with Positive Measure and a Space-Filling Curve, 124

Tong, Jingcheng and J. Chew, Some Remarks on Weak Continuity, 931

Toussaint, Godfried, Anthropomorphic Polygons, 31

Trimble, S. Y. and L.Pursell, Gram–Schmidt Orthogonalization by Gauss Elimination, 544

Villani, Alfonso, On the Equality $L^1(\mu) = L^1_{\text{loc}}(\mu)$, 826

Wagon, Stan and D. Flath, How to Pick Out the Integers in the Rationals: An Application of Number Theory to Logic, 812

Warren, Henry S., Jr., $3Q$; $1A$, 611

Weiss, Marysia T., An Early Introduction to Dynamics, 635

Wen, Liu, An Analytic Technique to prove Borel's Strong Law of Large Numbers, 146

Whittaker, James V., An Analytical Description of Some Simple Cases of Chaotic Behaviour, 489

Whittington, Keith, Crossing the Square, 833

Wilcox, Calvin H., The Cauchy Problem for the Wave Equation with Distribution Data: an Elementary Approach, 401

Witmer, Jeffrey A. and T. L. Moore, Statistics Within Departments of Mathematics at Liberal Arts Colleges, 431

Zetterström, Hans-Olov and B. Lindström, Borromean Circles are Impossible, 340

Zettl, A. and M. K. Kwong, Norm Inequalities for the Powers of a Matrix, 533

SUBJECT INDEX

This index uses the 1985 revision of the AMS's 1980 Mathematics Subject Classification scheme.

00-XX GENERAL

-01 Elementary exposition; textbooks
B. R. JOHNSON 427

04-XX SET THEORY

A25 Axiom of choice and related propositions (Zorn's lemma, etc.) J. LEWIN 353

05-XX COMBINATORICS

A15 Combinatorial enumeration problems, generating functions I. M. GESSEL 530 L. H. RIDDLE 613 K. B. BOGART 927

11-XX NUMBER THEORY

A07 Congruences; primitive roots; residue systems
M. A. BERGER, A. FELTZENBAUM, A. S. FRAENKEL, AND R. HOLZMAN 739

A99 None of the above, but in this section R. BEIGEL 332

D09 Quadratic and bilinear equations J. B. KELLY 527

J68 Rational approximation to algebraic numbers
R. GRANT 354

J70 Continued fractions and generalizations
N. RICHERT 133

N13 Primes in progressions A. FÄSSLER 518

S99 None of the above, but in this section
A. A. CUOCO 355

12-XX FIELD THEORY AND POLYNOMIALS

D05 Polynomials: factorization J. ISBELL 749

15-XX LINEAR AND MULTILINEAR ALGEBRA; MATRIX THEORY

-01 Elementary exposition; textbooks L. RUDOLPH 349 P. BINDING 850

A03 Vector spaces, linear dependence, rank K. P. S. BHASKARA RAO AND R. RAMACHANDRA RAO 127

A09 Matrix inversion, generalized inverses
R. A. LESLIE 732

A21 Canonical forms, reductions, classification
J. I. HALL 336

A23 Factorization of matrices L. PURSELL AND S. Y. TRIMBLE 544 D. DOKOVIĆ 935

A45 Miscellaneous inequalities involving matrices
M. K. KWONG AND A. ZETTL 533

A57 Other types of matrices (Hermitian, skew-Hermitian, etc.) H. LIEBECK AND A. OSBORNE 131 D. DE CAEN 829

A63 Quadratic and bilinear forms, inner products
P. BINDING 949

A69 Multilinear algebra, tensor products
N. METROPOLIS AND G.-C. ROTA 328

20-XX GROUP THEORY AND GENERALIZATIONS

-01 Elementary exposition; textbooks
J. D. REID 549

26-XX REAL FUNCTIONS

-01 Elementary exposition; textbooks
L. GILLMAN 346 D. R. CHALICE 255

A06 Functions of one variable; one-variable calculus
W. CARLIP 154 D. HARTIG 156 M. T. WEISS 635 H. ARMSTRONG AND R. HERZ-FISCHLER 752

A24 Functions of one variable; differentiation (functions of one variable): general theory, generalized derivatives, mean-value theorems
I. ROSENHOLTZ 149 S. KUHN 40 J. KLIPPERT 751 C. W. GROETSCH 847

A36 Antidifferentiation L. R. BRAGG 259

D15 Inequalities for sums, series, and integrals
K. B. STOLARSKY 889

30-XX FUNCTIONS OF A COMPLEX VARIABLE

-01 Elementary exposition; textbooks
A. GLUCHOFF 641

-02 Advanced exposition (research surveys, monographs, etc.) H. S. BEAR 109

E20 Integration, integrals of Cauchy type, integral representations of analytic functions P. A. LOEB 242

34-XX ORDINARY DIFFERENTIAL EQUATIONS

A30 Linear equations and systems C. C. ROSS, JR. 747

41-XX APPROXIMATIONS AND EXPANSIONS

A10 Approximation by polynomials J. H. LINDSEY, II 429

A15 Spline approximation F. B. RICHARDS AND J. FOSTER 47 C. DE BOOR AND K. HOLLIG 793

46-XX FUNCTIONAL ANALYSIS

-01 Elementary exposition; textbooks M. FURI AND M. MARTELLI 840

E30 Spaces of measurable functions(L^p-spaces, Orlicz spaces, Köthe function spaces, Lorentz spaces, rearrangement invariant spaces, etc.) A. G. MIAMEE 342

54-XX GENERAL TOPOLOGY

-01 Elementary exposition; textbooks D. GOFFINET 249

E52 Completions, Baire category D. JANKOVIC, M. GANSTER, AND I. REILLY 624

60-XX PROBABILITY THEORY AND STOCHASTIC PROCESSES

-01 Elementary exposition; textbooks T. L. MOORE AND J. A. WITMER 431

C05 Combinatorial probability S. G. PENRICE 617

F05 Central limit and other weak theorems L. TAKACS 742

F15 Strong theorems L. WEN 146

65-XX NUMERICAL ANALYSIS

F35 Numerical linear algebra; matrix norm conditioning, scaling Y. NIEVERGELT 539

90-XX ECONOMICS, OPERATIONS RESEARCH, PROGRAMMING, GAMES

C08 Special problems of linear programming (transportation and multi-index problems, etc.) A. R. MEIJER 144

REVIEWS BY TITLE

Names of authors are in ordinary type; those of reviewers in capitals.

Abstract and Concrete Categories, JiříAdámek, Horst Herrlich, and George Strecker, GABRIELE CASTELLINI, 285

Basic Hypergeometric Series, George Gasper and Mizan Rahman, GEORGE E. ANDREWS, 282

Beyond the Third Dimension, Banchoff, Thomas, H. S. M. COXETER, 668

Concrete Mathematics: A Foundation for Computer Science, Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, EDWARD A. BENDER, 779

Foundations for Advanced Mathematics, Carol Avelsgaard, JUDITH PALAGALLO, 179

Global Differential Geometry, S. S. Chern (ed.), THOMAS F. BANCHOFF, 669

How Does One Cut a Triangle?, Alexander Soifer, M. S. KLAMKIN, 775

Introduction to Advanced Mathematics, William Barnier and Norman Feldman, JUDITH PALAGALLO, 179

Japanese Temple Geometry Problems, Hidetosi Fukagawa and Dan Pedoe, DAN SOKOLOWSKY, 381

Mathematics and Its History, John Stillwell, JEFFREY NUNEMACHER, 569

The Mathematics of Projectiles in Sport, Neville de Mestre, VICTOR G. SZEBEHELY, 178

Pearls in Graph Theory, a Comprehensive Introduction, Nora Hartsfield and Gerhard Ringel, JOAN P. HUTCHINSON, 873

Recent Advances in Geometric Inequalities, D. S. Mitrinović, J. E. Pečarić, and V. Volonec, DAN PEDOE, 977

Representations and Characters of Finite Groups, M. J. Collins, MARK STEVEN MAZUR, 453

Symmetries of Culture: Theory and Practice of Plane Pattern Analysis, Dorothy K. Washburn and Donald W. Crowe, DORIS SCHATTSCHNEIDER, 383

Theorist, Prescience Corporation, FRANK WATTENBERG, 455

A Transition to Advanced Mathematics, Douglas Smith, Maurice Eggen and Richard St. Andre, JUDITH PALAGALLO, 179

REVIEWS BY AUTHOR

Names of authors are in ordinary type; those of reviewers in capitals.

Adámek, Jiří, H. Herrlich, and G. Strecker *Abstract and Concrete Categories,* GABRIELE CASTELLINI 285

Avelsgaard, Carol, *Foundations for Advanced Mathematics,* JUDITH PALAGALLO 179

Banchoff, Thomas, *Beyond the Third Dimension,* H. S. M. COXETER 668

Barnier, William and N. Feldman, *Introduction to Advanced Mathematics,* JUDITH PALAGALLO 179

Chern, S. S. (ed.), *Global Differential Geometry,* THOMAS F. BANCHOFF 669

Collins, M. J., *Representations and Characters of Finite Groups,* MARK STEVEN MAZUR 453

Crowe, Donald W. and D. K. Washburn, *Symmetries of Culture: Theory and Practice of Plane Pattern Analysis,* DORIS SCHATTSCHNEIDER 383

de Mestre, Neville, *The Mathematics of Projectiles in Sport,* VICTOR G. SZEBEHELY 178

Eggen, Maurice, D. Smith, and R. St. Andre *A Transition to Advanced Mathematics,* JUDITH PALAGALLO 179

Feldman, Norman and W. Barnier, *Introduction to Advanced Mathematics,* JUDITH PALAGALLO 179

Fukagawa, Hidetosi and D. Pedoe, *Japanese Temple Geometry Problems,* DAN SOKOLOWSKY 381

Gasper, George and M. Rahman, *Basic Hypergeometric Series,* GEORGE E. ANDREWS 282

Graham, Ronald L., D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science,* EDWARD A. BENDER 779

Hartsfield, Nora and G. Ringel, *Pearls in Graph Theory, a Comprehensive Introduction,* JOAN P. HUTCHINSON 873

Herrlich, Horst, J. Adámek, and G. Strecker *Abstract and Concrete Categories,* GABRIELE CASTELLINI 285

Knuth, Donald E., R. L. Graham, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science,* EDWARD A. BENDER 779

Mitrinović, D. S., J. E. Pečarić, and V. Volonec, *Recent Advances in Geometric Inequalities,* DAN PEDOE 977

Patashnik, Oren, R. L. Graham, and D. E. Knuth, *Concrete Mathematics: A Foundation for Computer Science,* EDWARD A. BENDER 779

Pečarić, J. E., D. S. Mitrinović, and V. Volonec, *Recent Advances in Geometric Inequalities,* DAN PEDOE 977

Pedoe, Dan and H. Fukagawa, *Japanese Temple Geometry Problems,* DAN SOKOLOWSKY 381

Prescience Corporation, *Theorist,* FRANK WATTENBERG 455

Rahman, Mizan and G. Gasper, *Basic Hypergeometric Series,* GEORGE E. ANDREWS 282

Ringel, Gerhard and N. Hartsfield, *Pearls in Graph Theory, a Comprehensive Introduction,* JOAN P. HUTCHINSON 873

Smith, Douglas, M. Eggen, and R. St. Andre, *A Transition to Advanced Mathematics,* JUDITH PALAGALLO 179

Soifer, Alexander, *How Does One Cut a Triangle?,* MURRAY S. KLAMKIN 775

St. Andre, Richard, D. Smith, and M. Eggen, *A Transition to Advanced Mathematics,* JUDITH PALAGALLO 179

Stillwell, John, *Mathematics and Its History,* JEFFREY NUNEMACHER 569

Strecker, George, J. Adámek, and H. Herrlich, *Abstract and Concrete Categories,* GABRIELE CASTELLINI 285

Volonec, V., D. S. Mitrinović, and J. E. Pečarić, *Recent Advances in Geometric Inequalities,* DAN PEDOE 977

Washburn, Dorothy K. and D. W. Crowe, *Symmetries of Culture: Theory and Practice of Plane Pattern Analysis,* DORIS SCHATTSCHNEIDER 383

INDEX OF SPECIAL TOPICS

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Shirley A. Hill 1

The Fiftieth William Lowell Putnam Competition 319

SOLUTIONS

Numbers in boldface refer to problems; those in lightface to pages.

E 3220 956
E 3257 55
E 3263 57
E 3274 958
E 3292 646
E 3293 59
E 3296 159
E 3299 60
E 3306 962
E 3306 264
E 3307 755
E 3308 60
E 3312 160
E 3315 366
E 3319 964
E 3323 162
E 3324 267
E 3326 62
E 3327 164
E 3328 367
E 3329 267
E 3330 760
E 3332 853
E 3333 553
E 3338 268
E 3339 269
E 3340 62
E 3341 165
E 3342 166
E 3343 368
E 3344 649
E 3345 439
E 3346 964
E 3347 167
E 3346 368
E 3348 855
E 3349 554
E 3350 555
E 3351 762
E 3352 369
E 3353 271
E 3354 440
E 3355 371
E 3356 371
E 3357 441
E 3358 650
E 3359 168
E 3361 443
E 3362 763
E 3364 444
E 3365 556
E 3367 960
E 3368 763
E 3369 764
E 3370 650
E 3371 652
E 3374 961
E 3375 652
E 3377 765
E 3380 557
E 3383 857
E 3384 961
E 3385 858
E 3387 653
E 3389 859
E 3391 860
E 3396 861
6506 169
6513 373
6568 375
6571 656
6574 171
6583 560
6587 660
6588 446
6591 64
6592 273
6593 65
6594 67
6595 377
6596 67
6597 862
6598 70
6599 767
6601 174
6602 176
6603 378
6604 865
6605 276
6606 277
6607 278
6608 966
6609 279
6610 448
6611 450
6612 663
6613 451
6614 561
6615 562
6618 565
6619 665
6620 769
6621 567
6622 770
6624 666
6626 771
6627 968
6628 772
6629 867
6630 868
6631 870
6636 970

PROBLEMS PROPOSED

Adrian, Constantin 754
Aoyagi, Tatsuhiko 365
Baker, John A. 646
Ball, Keith and Herman J. Tiersma 168
Baloglou, George and Phil Tracy 766
Bernau, S. J. and Gavin G. Gregory 852
Bishop, Richard L. and Lee A. Rubel 169
Bloom, D. M. 272
Boghossian, Artin B. 263
Borwein, Peter 553
Bowman, Douglas 553
Calbeck, William and Bruce Reznick 955
Callan, David 755
Cater, F. S. and John Erdman 559
Cavaretta, A. S. 646
Chenglong, Xu 438
Cooper, Curtis, R. E. Kennedy, and S. Rabinowitz 853
Demir, H. and C. Tezer 158
Demir, Huseyin 955
de Weger, B. M. M. 955
Doković, Dragomir Ž. 372, 767, 852
Egerland, W. O. and C. E. Hansen 273
Ehrhart, E. 755
Erdős, Paul 159, 264, 372, 655, 767, 965
Erdman, John and F. S. Cater 559
Feil, Todd and Gary Kennedy 853
Fenton, William 852
Flajolet, Philippe and Donald E. Knuth 54
Friedman, Jane 437
Gülicher, Herbert 366
Gillman, Richard Alan 446
Goffinet, Daniel 63
Golomb, Michael 366
Gregory, Gavin G. and S. J. Bernau 852
Grivaux, Jean-Pierre 552
Hansen, C. E. and W. O. Egerland 273

Horwitz, Alan 158
Hoyt, John P. 365
Jónsson, Wilbur and Gérard Letac 438
Janous, Walther 158, 559
Jansen, Ronald A. and Pieter Moree 552
Jeurissen, R. H. 862
Kennedy, Gary and Todd Feil 853
Kennedy, Robert E., C. Cooper, and S. Rabinowitz 853
Khan, M. A. 645
Knuth, Donald E. 168, 852
Knuth, Donald E. and John McCarthy 264
Knuth, Donald E. and Philippe Flajolet 54
Kuczma, Marcin E. 55
Kushner, H. B. 862
Lagarias, Jeffrey C. and Thomas Zaslavsky 559
Lee, D. K. and B. A. Murray 273
Lenard, Andrew 956
Letac, Gérard and Wilbur Jónsson 438
Luthar, R. S. 54
Malde, Paresh J. and Allen J. Schwenk 965
McCarthy, John and Donald E. Knuth 264
Mendelsohn, N., R. Padmanabhan, and B. Wolk 263
Merzlyakov, S. G. 55
Moree, Pieter and Ronald A. Jansen 552
Murray, B. A. and D. K. Lee 273
Neuts, Marcel F. 655
Nicol, C. A. and J. L. Selfridge 645
Padmanabhan, R., N. S. Mendelsohn, and B. Wolk 263
Palacios, José Luis and Dennis P. Sandell 655
Parker, E. T. 55
Pinsky, Mark A. 553
Pomerance, Carl 862
Popescu, Călin 438
Räde, Lennart 366
Rabinowitz, Stanley, C. Cooper, and R.E. Kennedy 853
Reznick, Bruce and William Calbeck 955
Rogers, D. G. 646
Rosset, Shmuel and Tamir Shalom 159
Rotman, J. J. 755
Rotman, J. J. and P. M. Weichsel 372
Rubel, Lee A. and Richard L. Bishop 169
Rudin, Walter 63
Sandell, Dennis and José Luis Palacios 655
Schwenk, Allen J. 645
Schwenk, Allen J. and Paresh J. Malde 965
Selfridge, J. L. and C. A. Nicol 645
Shallit, Jeffrey 264
Shalom, Tamir and Shmuel Rosset 159
Snevily, Hunter 956
Steele, J. Michael 159
Tóth, László 264
Taylor, Howard 553
Tezer, C. and H. Demir 158
Tiersma, Herman J. and Keith Ball 168
Tissier, Alain 965
Tracy, Phil and George Baloglou 766
Ungar, Abraham 445
Van Hamme, L. 445
Vanden Eynden, Charles 365
Vince, A. J. 63
Wallen, Lawrence J. 956
Wardlaw, William P. 437
Weichsel, P. M. and J. J. Rotman 372
Wilf, Herbert S. 754
Wolk, B., R. Padmanabhan, and N. S. Mendelsohn 263
Wylie, Ray 438
Zannier, Umberto 754
Zaslavsky, Thomas and Jeffrey C. lagarias 559
Zeifman, A. 54

PROBLEMS SOLVED

Adin, Ron 652
Alexopoulos, John 450
Andersen, K. F. 441, 854
Andrews, Peter, M. S. Klamkin, and E. T. H. Wang 755
Anglesio, Jean 265, 861
Apostol, Tom M. 563
Bailey, Duane W. 761
Bang, Seung-Jin 369
Behler, Kerry 167
Bialostocki, Arie and J. Schönheim 653
Boese, G. 968
Borwein, David 276
Botsko, Michael W. 267
Bowman, Doug and Tad White 279
Boyd, David W. 451
Bremser, Priscilla S. 660
Briggs, W. E. 867
Brison, Owen and Jorge-Nuno O. Silva 61
Buddenhagen, Jim 773
Buhler, Wolfgang 853
Callan, David 442, 961, 963
Callan, David and Douglas West 960
Callan, David and Grace Whaba 64
Chapman, R. J. 556, 665
Chapman, R. J. and R. W. K. Odoni 269
Chico Problem Group 371
Cobb, John 70
Curtin, Eugene 663
Dalang, Robert C. 770
de Weger, B. M. M. 174, 272
Duemmel, James and Allan Mauney 445
Elkies, Noam D. 865
Estes, Dennis R. 567
Evans, Ronald J. 865

Evans, Ronald and Montgomery, Peter 870
Falkowitz, Mordechai 268
Ferrer, Jesús 62
Fine, Nathan J. 860
Ford, Kevin 563
Foster, L. L. 443
Galvin, Fred and Robert High 666
Grant, Christopher P. 65
Gritzman, P., B. Mohar, J. Pach, and R. Pollack 165
Grivaux, Jean-Pierre 267
Gross, Benedict H. 169
Grossman, Jerrold W. 554
Guy, Richard K. 59
Herzog, Joachim, Kee-Wai Lau 446
Heuer, Charles V. 647
High, Robert and Fred Galvin 666
Isaacs, I. Martin 277
Ismail, Mourad E. H. 278
Ivić, Aleksander 767
Köditz, H, L. Mattner, E. Mues, and M. Reinders 448
Kappus, Hans 60
Keane, Joseph and Gregg Patruno 59
Kløve, Torleiv 369
Klamkin, M. S., Peter Andrews, and E. T. H. Wang 755
Kleitman, Daniel J. 958
Kozma, Ilan 164
Kresch, Andrew 651
Kronheimer, Erwin 956
Kuczma, Marcin E. 267, 858, 859
Lau, Kee-Wai 271, 553
Lau, Kee-Wai and Joachim Herzog 446
Levine, Eugene 273
Lindsey, J. H., II 970
Lossers, O. P. 168, 764, 957, 962
Machover, Maurice 373
Maderer, Charles R. 773
Martin, Reiner 271, 442
Mattics, L. E. 171, 560
Mattner, L., H. Köditz, E. Mues, and M. Reinders 448
Mauldon, J. G. 55
Mauney, Allen and James Duemmel 445
Merickel, James G. 557
Mimura, Yoshio 378
Min, Woo Jung 58
Mohar, B., P. Gritzman, J. Pach, and R. Pollack 165
Mohler, Lee 771
Montgomery, Peter and Ronald Evans 870
Mues, E., H. Köditz, L. Mattner, and M. Reinders 448
Munkries, James R. 771
Newcomb, William A. 656
Nieto, José Heber 166, 366
Noland, Hugh 371
Odoni, R. W. K. and R. J. Chapman 269
Pach, J., P. Gritzman, B. Mohar, and R. Pollack 165
Pedersen, Allan 766, 863
Pierce, Stephen 763
Pollack, R., P. Gritzman, B. Mohar, and J. Pach 165
Potters, J. A. M. 58
Ranaldi, Paolo and Carl Schoen 368
Reinders, M., H. Köditz, L. Mattner, and E. Mues 448
Richman, Fred 377
Rickard, J. 957
Riese, Adam 769
Robinson, Raphael M. 367
Ross, David S. 857
Sarkar, Amites 653
Sasvári, Zoltán 966
Schönheim, J. and Arie Bialostocki 653
Schilling, Kenneth 440, 868
Schoen, Carl and Paolo Ranaldi 368
Schwenk, Allan J. 159
Secrest, David 176
Shore, Terence and Douglas B. Tyler 764
Silva, Jorge-Nuno O. and Owen Brison 61
Simpson, R. J. 443
Spencer, Philip and Richard Stong 651
Steelman, John Henry 63, 371, 439, 555, 565, 764, 773
Stong, Richard 162, 168, 267, 371, 555, 561, 649, 762, 855, 861, 970
Stong, Richard and Philip Spencer 651
Stoops, Glenn A. 853
Sumner, John S. 264
Tyler, Douglas B. and Terence Shore 764
Ullman, Daniel 168
University of South Alabama Problem Group 375
Vowe, Michael 265
WMC Problems Group 650
Wahba, Grace and David Callan 64
Want, E. T. H., Peter Andrews, and M. S. Klamkin 755
Wend, David V. V. 67
West, Douglas and David Callan 960
White, Tad and Doug Bowman 279

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 98, Number 10 December 1991

Contents

(ISSN 0002-9890)

ARTICLES

From Wythoff's Nim to Chebyshev's Inequality KENNETH B. STOLARSKY 889

A Topological Approach to Digital Topology T. YUNG KONG, RALPH KOPPERMAN, AND PAUL R. MEYER 901

Linear Algebra by Analogy SCOTT H. HOCHWALD 918

NOTES

An Obvious Proof of Burnside's Lemma KENNETH B. BOGART 927

Poisson Integrals of Riemann Integrable Functions J. KRÁL AND W. F. PFEFFER 929

Some Remarks on Weak Continuity JAMES CHEW AND JINGCHENG TONG 931

On the Product of Two Alternating Matrices DRAGOMIR Ž. DOKOVIĆ 935

THE TEACHING OF MATHEMATICS

Reliability, Recursion, and Risk LAVON B. PAGE AND JO ELLEN PERRY 937

An "Ocean of Zeros" Proof That a Certain Non-Liouville Number is Transcendental M. J. KNIGHT 947

Simple Tests for Classifying Critical Points of Quadratics with Linear Constraints PAUL BINDING 949

PROBLEMS AND SOLUTIONS

Elementary Problems and Solution 955

Advanced Problems and Solutions 965

Monthly Unsolved Problems, 1969–1991 RICHARD K. GUY 973

REVIEWS

Recent Advances in Geometric Inequalities. By D. S. Mitrinović, J. E. Pečarić, and V. Volonec DAN PEDOE 977

IN APPRECIATION 981

TELEGRAPHIC REVIEWS 983

INDEX TO VOLUME 98 989

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts, including Notes and longer articles, should be sent to the editor-elect:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Proposed problems or solutions should be sent to

RICHARD BUMBY
P.O. Box 10971
New Brunswick, NJ 08906

Please send 3 copies of all material.

Letters to the Editor, both for publication and for private reading, should be sent to the editor-elect at the address given above. We welcome suggestions for new ways of making the MONTHLY a lively, entertaining, and informative journal.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $33 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 34% dues discount for the first two years of membership.) The nonmember/library subscription price is $128 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1991, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

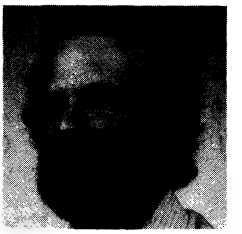
Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

From Wythoff's Nim to Chebyshev's Inequality

KENNETH B. STOLARSKY,* *University of Illinois, Urbana*



KENNETH B. STOLARSKY received his undergraduate education at Caltech, and obtained a Ph.D. from the University of Wisconsin (Madison) in 1968 under the supervision of Marvin I. Knopp. He did one year of postdoctoral work at the Institute for Advanced Study (Princeton), and has been (aside from sabbaticals) at the University of Illinois (Champaign-Urbana) since then, where he is now professor of mathematics. In 1979, 1980 and 1981 he was a member of the national Putnam Contest Committee; for the past five years he has served as an Associate Editor of this Monthly. His research interests include number theory, combinatorics, extremal problems of geometry, and complex analysis (with emphasis on polynomials).

1. Introduction. This is the story of how the study of a combinatorial game, Wythoff's Nim, led to certain inequalities for real integrals. Since J. H. Conway [**C**; see also **B-C-G**] has shown that the construction of the real numbers (and rather more) can be based on a collection of combinatorial games, this is not totally implausible. However, our point of view is of a different nature than Conway's. We consider only Wythoff's Nim, and from a purely algebraic point of view. This leads to questions about the associativity of certain multiplications. A probabilistic investigation then suggests inequalities involving certain multiple integrals. We first prove these in a spirit consonant with their original context, but then discover that Chebyshev's (Tchebychef's) inequality for similarly ordered sequences can be taken as the underlying principle. Our journey from associative multiplications for the integers to inequalities for integrals ends in a curious way. The inequalities themselves suggest an uncountable collection of associative multiplications for the positive real numbers (§9).

Wythoff's Nim (or simply Wythoff's game) is described in many references, for example [**Ba-C, B, B-C-G, Co, Cox, F, F-B, G, Wy**]; see also many of the references cited in [**S1**]. The initial positions of the game are pairs of nonnegative integers (P, Q). Play alternates between two players. At each turn the player's move consists of replacing (P, Q) by one of

$$(P - M, Q), (P, Q - M), \quad \text{or} \quad (P - M, Q - M)$$

where (i) $M \geq 1$ is an integer and (ii) the new pair is also a pair of nonnegative integers. The first player unable to move loses. Wythoff proved [**Wy**] that if neither player blunders, the player about to move will win if and only if he can move to a pair (P', Q') such that either it or its transpose (Q', P') has the form

$$(a(n), b(n)) = \left([n\phi], [n\phi^2]\right).$$

Here $[u]$ denotes the greatest integer in u and $\phi = 1.618\ldots$ is the "golden mean," i.e. the largest root of $x^2 - x - 1 = 0$. Such pairs are called (Wythoff) winning pairs. The first six are $(0, 0)$, $(1, 2)$, $(3, 5)$, $(4, 7)$, $(6, 10)$, and $(8, 13)$.

*Supported by ONR Grant N00014-85-K-0368

2. Certain inequalities for integrals. Our principal inequality can be stated very simply.

THEOREM. *If* $0 \le g(u) \le 1$ *and* $g(u)$ *is nonincreasing on* $[0, 1]$, *then for positive* n *and* m *we have*

$$\int_0^1 g(u^{1/n+m})\,du \ge \int_0^1 g(u^{1/n})\,du \int_0^1 g(u^{1/m})\,du. \tag{2.1}$$

However it seems easier to work with (and "get a feeling for") a rather less concise formulation of this Theorem. If $g(u)$ is integrable on $[0, 1]$ and $n > 0$, we define the quotient $Q = Q(g; n)$ by

$$Q = Q(g; n) = \frac{\int_0^1 g(u)u^{n-1}\,du}{\int_0^1 u^{n-1}\,du} \tag{2.2}$$

whenever the ratio on the right exists. This is a weighted average of g.

THEOREM. *If* $0 \le g(u) \le 1$ *and* $g(u)$ *is nonincreasing on* $[0, 1]$, *then for positive* n *and* m *we have*

$$Q(g; n + m) \ge Q(g; n)Q(g; m). \tag{2.3}$$

Equality holds for commensurable m *and* n *if and only if there is a real* c *with* $0 \le c \le 1$ *such that*

$$g(u) = \begin{cases} 1 & 0 \le u < c \\ 0 & c < u \le 1. \end{cases} \tag{2.4}$$

The original formulation is easily recovered from (2.3) by three simple changes of variable. It seems natural to conjecture that the above equality criterion holds for all positive real m and n. The author is embarrassed by his inability to prove this; see, however, the Proposition in §7.

For g nondecreasing a similar Theorem is valid with u^{n-1} replaced by $(1 - u)^{n-1}$ and the functional values 0 and 1 interchanged in the criterion for equality.

Our main point is not some possible novelty in (2.1) *but rather the manner in which it arose* (*see* §3), *was proved* (*see* §4 *and* §5), *and served to suggest other inequalities* (*see* §6 *and* §8). We shall give a fairly straightforward proof of it based on Chebyshev's inequality in §7.

3. The semigroup of Wythoff's game. It was recently discovered [**P-S2**] that the winning positions of Wythoff's game form a multiplicative semigroup. Associate to each winning pair $(a(n), b(n))$ the real number

$$W(n) = b(n) - \phi a(n). \tag{3.1}$$

This is a one-to-one map of the pairs into $[0, 1]$. These numbers are closed under multiplication! In fact,

$$W(n) = \{n\phi\}/\phi, \quad \text{and} \quad W(n)W(m) = W(n * m) \tag{3.2}$$

where $\{x\}$ denotes the fractional part of x and

$$n * m = nm + a(n)a(m). \tag{3.3}$$

The first statement in (3.2) is easily verified. For the proof of the second assertion we use the following Lemma (the "*K-L-M* formula").

LEMMA. *For any integers K, L, M and n we have*

$$a(Ka(n) + Ln + M) = Kb(n) + La(n) + \left\lfloor M\phi + (L\phi - K)\frac{\{n\phi\}}{\phi} \right\rfloor. \quad (3.4)$$

Proof. Let m be a variable integer. Then (here and elsewhere the notation $z:=$ means that the new symbol nearest the : is being defined)

$$\begin{aligned}\Delta:&= (Ka(n) + Ln + M)\phi - (Kb(n) + La(n) + m)\\ &= L(n\phi - a(n)) - K(b(n) - a(n)\phi) + M\phi - m\\ &= L\{n\phi\} - K\frac{\{n\phi\}}{\phi} + M\phi - m\\ &= \left((L\phi - K)\frac{\{n\phi\}}{\phi} + M\phi\right) - m = :R - m.\end{aligned}$$

Now choose $m = [R]$. Then $0 \le \Delta < 1$ and the Lemma follows.

Now, from the *K-L-M* formula and the fact that $b(m) = a(m) + m$, we have

$$\begin{aligned}a(a(m)a(n) + mn) &= a(m)b(n) + ma(n) + \left\lfloor (m\phi - a(m))\frac{\{n\phi\}}{\phi} \right\rfloor\\ &= (b(m) - m)b(n) + m(b(n) - n)\\ &= b(m)b(n) - mn,\end{aligned}$$

so

$$b(a(m)a(n) + mn) = b(m)b(n) - mn + mn + a(m)a(n)$$

and

$$\begin{aligned}-a(a(m)a(n) + mn) &= mn - b(m)b(n)\\ &= -na(m) - ma(n) - a(n)a(m)\\ &= a(n)a(m) - a(n)b(m) - a(m)b(n).\end{aligned}$$

Thus

$$\begin{aligned}W(n)W(m) &= b(m)b(n) + a(m)a(n) + \phi(a(m)a(n)\\ &\qquad - a(n)b(m) - a(m)b(n))\\ &= b(n*m) - \phi a(n*m)\end{aligned}$$

and multiplicative closure is established.

Since ordinary multiplication is associative, we have

$$(W(n)W(m))W(k) = W(n)(W(m)W(k)) \quad (3.5)$$

and hence the $*$-multiplication is *associative*. Do some similar multiplication rules also have the associativity property? Indeed, a considerable number of such rules have now been found [**Ar**, **K**, **P-S2**, **P-S3**]. One of the most general ways (so far) of finding such rules has been to study [**P-S2**] the set S of all integral affine transformations that map the set of all Wythoff pairs $(a(n), b(n))$ back into itself. This set S is a two-parameter noncommutative semigroup isomorphic to the set of

all pairs of integers (j, l) with the associative (yes!) multiplication rule

$$(j,l) \otimes (k,m) = \big(t_{jl}(k), t_{jl}(m)\big) \tag{3.6}$$

where

$$t_{jl}(n) = (j-l)a(n) + (c(j) - c(l))n + j \tag{3.7}$$

and $c(j) = a(j) - j$ for all j. Quite a few of the associative rules so far discovered correspond to subsemigroups of S. A. Fraenkel, H. Porta, and the author (unpublished work) have also found a rich variety of associative laws connected with quadratic irrationalities other than the golden mean ϕ.

One seemingly slight variation of $n * m$ not investigated in these references is

$$n\#m = 4nm + a(2n)a(2m). \tag{3.8}$$

It is tempting to say this is associative since we are simply using $2n$ as a "code" for n. A closer look, however, reveals no strong reason for the associative law to hold. Several initial attempts to produce a counterexample by hand (with n and m chosen at random) failed, and my first counterexample,

$$(8\#42)\#110 \neq 8\#(42\#110), \tag{3.9}$$

was obtained by choosing the numbers "maliciously". The difficulty of finding a counterexample calls for an explanation. It turns out that if (k, m, n) is a lattice point chosen "at random" from the positive octant, then the probability of

$$(k\#m)\#n = k\#(m\#n) \tag{3.10}$$

exceeds 96%! (It is clear what "at random" means if (k, m, n) is chosen from

$$\{(k_1, k_2, k_3) : 0 \leq k_i \leq N,\ 1 \leq i \leq 3, \text{ and } k_i \text{ is integral}\}. \tag{3.11}$$

We define the "at random" probability to be the limit as $N \to \infty$.)

The key to computing the probability is the following identity (a consequence of the *K-L-M* formula):

$$\begin{aligned} k\#(m\#n) = {} & 16kmn + 4(ka(2m)a(2n) + ma(2n)a(2k) + na(2k)a(2m)) \\ & + 2a(2k)a(2m)a(2n) + a(2k)\big[2\phi^{-1}\{2m\phi\}\{2n\phi\}\big], \end{aligned} \tag{3.12}$$

where the [] denotes "greatest integer in". We also need the very well-known result [**Cas**, **K-N**] that if a is irrational then the sequence $\{na\}$, $n = 1, 2, 3, \ldots$, is uniformly distributed in $[0, 1]$.

4. A probability integral. By the last identity of §3, the associativity statement

$$k\#(m\#n) = (k\#m)\#n \tag{4.1}$$

is equivalent to

$$a(2k)\left[\frac{2}{\phi}\{2m\phi\}\{2n\phi\}\right] = a(2m)\left[\frac{2}{\phi}\{2k\phi\}\{2n\phi\}\right]. \tag{4.2}$$

If $m = k$ (an event of probability zero) this is clearly true. Note that the greatest integer terms can only be 0 or 1. Set $x = \{2k\phi\}$, $y = \{2m\phi\}$, and $z = \{2n\phi\}$. It is easy to see that equality holds for $m \neq k$ if and only if

$$xz < \phi/2 \quad \text{and} \quad yz < \phi/2. \tag{4.3}$$

The "individual" probabilities are easy to calculate, but these events are not independent. *In fact, the occurrence of one of them clearly makes the other more probable.* This observation is enough to establish that the probability exceeds .961333... since

$$\int_0^1 \min\left(1, \frac{\phi}{2x}\right) dx = \int_0^{\phi/2} dx + \int_{\phi/2}^1 \frac{\phi}{2x}\, dx$$

$$= \frac{\phi}{2} - \frac{\phi}{2} \ln \frac{\phi}{2}$$

$$= .9804759\ldots = \sqrt{.961333\ldots}$$

To find the exact probability we shall need to calculate a multiple integral.

LEMMA. *For* $0 \le b \le 1$ *we have*

$$I_n = \int_0^1 \cdots \int_0^1 \min\left(1, \frac{b}{x_1}, \ldots, \frac{b}{x_n}\right) dx_1 \cdots dx_n$$

$$= \begin{cases} \dfrac{nb - b^n}{n-1}, & n \ge 2 \\ b - b \ln b, & n = 1. \end{cases} \tag{4.4}$$

Proof. This is easy for $n = 1$. The idea for $n \ge 2$ is to partition the unit cube of E^n into the $n!$ simplices given by the $n!$ different possible orderings of the variables $x_1, \ldots, x_n$. On each of these the integral of the integrand will have the same value, namely

$$\int_0^1 \int_0^{x_n} \cdots \int_0^{x_3} \int_0^{x_2} \min\left(1, \frac{b}{x_n}\right) dx_1 \cdots dx_n$$

$$= \int_0^1 \frac{x_n^{n-1}}{(n-1)!} \min\left(1, \frac{b}{x_n}\right) dx_n$$

$$= \int_0^b \frac{x^{n-1}}{(n-1)!}\, dx + b \int_b^1 \frac{x^{n-2}}{(n-1)!} = \frac{1}{n!} \frac{nb - b^n}{n-1}, \tag{4.5}$$

and the result follows.

We can now determine the probability of equality in (4.1). Since 2ϕ is irrational, the numbers x, y and z are random in $[0, 1]$. Thus

$$\text{Prob}[0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1, z < \phi/2x, \text{ and } z < \phi/2y]$$

$$= \int_0^1 \int_0^1 \int_0^{\min(1, \phi/2x, \phi/2y)} dz\, dx\, dy$$

$$= I_2 = \frac{\phi}{2}\left(2 - \frac{\phi}{2}\right) = .963525\ldots. \tag{4.6}$$

It's rather close to the rough estimate! At this point the study of the integrals can be detached from the question that called them into being. If s is real the same

method of proof yields

$$\int_0^1 \cdots \int_0^1 \min\left(1, \left(\frac{b}{x_1}\right)^s, \ldots, \left(\frac{b}{x_n}\right)^s\right) dx_1 \cdots dx_n$$

$$= \begin{cases} 1 + sb^n(n-s)^{-1} & n \neq s,\ s \leq 0, \\ \dfrac{nb^s - sb^n}{n-s} & n \neq s,\ s \geq 0, \\ b^s - b^s \ln b^s & n = s. \end{cases} \tag{4.7}$$

That (for s a positive integer) the integral is symmetrical in n and s is far from obvious (at least to the author).

5. Inequalities for integrals. The italicized observation of §4 says that

$$P(A \cap B) = P(A)P(B|A) \geq P(A)P(B) \tag{5.1}$$

for certain special events A and B. We now drop the language of probability, but the above is the real motivation for what follows.

We can now prove the Theorem. Assume that n and m are positive integers. Let $N = n + m$ and let $t_1, \ldots, t_N$ be numbers in $[0, 1]$. Clearly

$$\min(t_1, \ldots, t_m, t_{m+1}, \ldots, t_{m+n}) \geq \min(t_1, \ldots, t_m) \cdot \min(t_{m+1}, \ldots, t_{m+n}). \tag{5.2}$$

Now let g be a nonincreasing function that maps $[0, 1]$ into itself. (For example, if $s \geq 0$ then

$$g = g_s(x) = \min\left(1, \left(\frac{b}{x}\right)^s\right)$$

is such a function). Then

$$\int_0^1 \cdots \int_0^1 \min(g(x_1), \ldots, g(x_N))\, dx_1 \cdots dx_N$$

$$= N! \int_0^1 \int_0^{x_N} \cdots \int_0^{x_2} g(x_N)\, dx_1 \cdots dx_N$$

$$= \frac{N!}{(N-1)!} \int_0^1 g(x_N) x_N^{N-1}\, dx_N = N \int_0^1 g(x) x^{N-1}\, dx$$

$$= Q(g; N) = Q(g; n + m). \tag{5.3}$$

The inequality now follows: simply minorize the integrand of (5.3) by the product on the right of (5.2). For the case of equality, first assume that g is neither 0 nor 1 on some open interval I. Then (5.2) holds with strict inequality for $(t_1, \ldots, t_N)$ in $I \times \cdots \times I$ (N times), so there is strict inequality in (2.3). Thus, since g is nonincreasing, equality implies that there is some c in $[0, 1]$ such that (except perhaps at c itself) g is 1 to the left of c and 0 to the right. Thus we have (2.4) and equality since

$$c^{n+m} = c^n c^m.$$

This proves the theorem for integral values of m and n. Now let D be positive and let t be a real variable varying between 0 and 1. The change of variable $u = t^{1/D}$

yields

$$n\int_0^1 g(u)u^{n-1}\,du = \frac{n}{D}\int_0^1 g(t^{1/D})t^{-1+n/D}\,dt.$$

As $g(u)$ varies over all nonincreasing functions on $[0, 1]$, so does $g(t^{1/D})$. Hence the inequality is true for n and m commensurable, and by taking limits it is easy to see that it is true for all positive real numbers n and m. The inverse change of variable $t = u^D$ enables us to show that the criterion for equality in the case where n and m are commensurable is the same as that in the integral case. This proves the Theorem.

6. Fundamental inequalities. Depending upon the ranges of x and q there are various "fundamental" inequalities between x and its "q-analogue", namely

$$x_q = \frac{q^x - 1}{q - 1}. \tag{6.1}$$

The q-analogue is so called because its limit as $q \to 1$ is x. It is stressed in [**B-B**, **H-L-P**] that these fundamental inequalities can be used to derive the most important inequalities of classical analysis, those of Hölder and Minkowski, and the inequality between the arithmetic and geometric means. (These q-analogue inequalities can also be interpreted from a convection-diffusion point of view [**S2**].) Observe that if $0 \le q \le 1$ (resp. $1 < q$) the inequality between the arithmetic and geometric means of q and 1, with weights x and $1 - x$, $0 \le x \le 1$, namely

$$q^x 1^{1-x} \le xq + (1 - x)1, \tag{6.2}$$

is

$$x \le x_q \,(\text{resp. } x_q \le x). \tag{6.3}$$

The obvious fact that the multiple integral of (4.7) is at most 1 yields for $0 \le n \le s$ (resp. $0 \le s \le n$) the inequality

$$\frac{n}{s} \le \frac{b^n - 1}{b^s - 1} \,(\text{resp. } \ge). \tag{6.4}$$

Upon making the substitutions $b \to q^{1/s}$ and $n/s \to x$ we obtain, with s a positive integer, the inequality (6.3) for rational x. The inequality for x real is a limiting case. Note that among these results are inequalities valid for $x \ge 1$. We omit a precise analysis of the cases of equality.

Now assume $0 \le b \le 1$ and apply the Theorem (see (4.7)) to the function

$$f(b, s; n) = \frac{nb^s - sb^n}{n - s} \tag{6.5}$$

We obtain the

COROLLARY. *For $0 \le b \le 1$, $s \ge 0$, and positive integers n and m, we have*

$$f(b, s; n + m) \ge f(b, s; n)f(b, s; m). \tag{6.6}$$

It seems strange that such a simple function could have a "previously undiscovered" property, but the above is new, at least to the author. A proof without

integrals is called for. Some manipulation shows it is equivalent to

$$F(b) = (m-s)(n-s)(n+m) - nmsb^{n+m-s} - nm(n+m-s)b^s + ns(n+m-s)b^m + ms(n+m-s)b^n \geq 0. \tag{6.7}$$

For $b = 0$ this is true if also $n + m \geq 0$, while for $b = 1$ we have equality. Hence it suffices to show $F(b)$ is decreasing. This is true if also $nm(n + m - s) \geq 0$ since then

$$F'(b) = -nm(n+m-s)b^{s-1}(1-b^{n-s})(1-b^{m-s}) \leq 0. \tag{6.8}$$

In particular, the Corollary remains true if n and m are both positive real numbers that exceed s.

The Corollary suggests some second order improvements (similar in spirit to those given in [**S2**]) in the inequalities between x and x_q for certain ranges of the parameters, but we shall not pursue this here.

7. Chebyshev's inequality. Consider the step function

$$g(u) = g_i, \qquad c_i \leq u < c_{i+1} \tag{7.1}$$

where $0 \leq i \leq N$, $c_0 = 0$, $c_{N+1} = 1$, $c_i < c_{i+1}$, $g_0 = 1$, $g_N = 0$, and $g_{i+1} \leq g_i$. A simple calculation shows that

$$p\int_0^1 g(u)u^{p-1}\,du = (g_0 - g_1)c_1^p + \cdots + (g_{N-1} - g_N)c_N^p = \sum_{i=1}^{N} \lambda_i c_i^p \tag{7.2}$$

where $\lambda_i \geq 0$ and $\sum\lambda_i = g_0 - g_N = 1$. Also, for $\varepsilon > 0$ and any bounded decreasing function f defined on $[0, 1]$ there is clearly such a step function g with

$$\int_0^1 |f - g|\,du < \varepsilon, \tag{7.3}$$

so it suffices to prove our inequality for such g. Now for $n, m > 0$ we have by direct expansion

$$0 \leq \sum_{j=1}^{N}\sum_{i=1}^{N} \lambda_i\lambda_j(c_i^n - c_j^n)(c_i^m - c_j^m) = 2\sum_{i=1}^{N} \lambda_i c_i^{n+m} - 2\sum_{i=1}^{N} \lambda_i c_i^n \sum_{j=1}^{N} \lambda_j c_j^m \tag{7.4}$$

and hence by (7.2) the validity of our inequality (2.3) for step functions. Now (7.4) is almost Chebyshev's inequality; if we replace the finite sequences $\{c_i^n\}$ and $\{c_i^m\}$ by sequences $\{a_i\}$ and $\{b_i\}$ such that

$$(a_i - a_j)(b_i - b_j) \geq 0 \tag{7.5}$$

for all i and j, and expand the corresponding double sum in (7.4), we obtain

$$\sum\lambda_i a_i \sum\lambda_i b_i \leq \sum\lambda_i a_i b_i. \tag{7.6}$$

This is Chebyshev's inequality [**H-L-P**, pp. 43–44]. Clearly there is strict inequality

unless either all the a_i are equal, or all the b_i are equal. In our own special case this can only happen when $N = 1$. Thus we have the

PROPOSITION. *If g is a step function, we have equality in* (2.3) *if and only if* (2.4) *holds.*

Remarks. Hardy-Littlewood-Polya omit any discussion of equality in their formulation of the integral analogue of Chebyshev's inequality [**H-L-P**, Theorem 236, p. 168].

Given the statement of the Theorem, it seems natural to study the product of the "n-integral" and the "m-integral" when $T = n + m$ is held constant. To do this it is natural to introduce the function

$$F(t) = \int_0^1 g(u)u^{T/2+t}\,du \int_0^1 g(u)u^{T/2-t}\,du. \tag{7.7}$$

Our theorem (together with the Cauchy-Schwarz inequality) becomes

$$\left(\int_0^1 g(u)u^{T/2}\,du\right)^2 \le F(t) \le \left(\frac{1}{\frac{T}{2}+t+1} + \frac{1}{\frac{T}{2}-t+1}\right)\int_0^1 g(u)u^{T+1}\,du; \tag{7.8}$$

for example ($t = 0$)

$$\left(\int_0^1 g(u)u^{T/2}\,du\right)^2 \le \frac{4}{T+2}\int_0^1 g(u)u^{T+1}\,du. \tag{7.9}$$

Note that the upper bound in (7.8) is a convex function of t. In fact (we omit the proof) it can be shown that $F(t)$ itself is convex for $|t| \le T/2$.

8. A Gamma function inequality. It seems appropriate to indicate what our inequality says about Beta and Gamma functions, especially since the Beta function integral appears in the next section. For an exceptional account of the role played by the Beta function in the general theory of special functions, see [**Ca**].

Let g be a nondecreasing function, so

$$F(n+m) \ge F(n)F(m) \tag{8.1}$$

where

$$F(n) = n\int_0^1 g(x)(1-x)^{n-1}\,dx \tag{8.2}$$

(recall the second paragraph after (2.4)).

If we take $g(x) = x^{a-1}$ where $a \ge 1$, then $F(n)/n$ becomes $B(a, n)$, the Beta function of a and n. Since

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{8.3}$$

where $\Gamma(x)$ is the gamma function, we have from (8.1) that

$$(m+n)B(a, m+n) \ge nB(a, n) \cdot mB(a, m). \tag{8.4}$$

Then from $x\Gamma(x) = \Gamma(x+1)$ we have that

$$\rho(a) := \frac{\Gamma(m+n+a)\Gamma(a)}{\Gamma(m+a)\Gamma(n+a)} \leq \frac{\Gamma(m+n+1)\Gamma(1)}{\Gamma(m+1)\Gamma(n+1)} \tag{8.5}$$

for $a \geq 1$. This *suggests* that in fact $\rho(a)$ is decreasing in a. Now Gauss' ${}_2F_1$ summation states that

$$\begin{aligned} {}_2F_1(a,b;c;1) &= 1 + \frac{ab}{c}\frac{1}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{1}{2!} + \cdots \\ &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \end{aligned} \tag{8.6}$$

for a, b, c complex, c not a negative integer, and (in case neither a nor b is a nonpositive integer)

$$\mathrm{Re}(c-a-b) > 0;$$

see, e.g. [**Ca**, **W-W**, pp. 281–282]. Thus

$$\frac{\Gamma(m+n+a)\Gamma(a)}{\Gamma(n+a)\Gamma(m+a)} = {}_2F_1(n,m;n+m+a;1)$$

and the infinite series makes it clear that this is decreasing in a.

9. Associative multiplications for real numbers. If we could ignore the greatest integer brackets implicit in the definition of $n * m$ in (3.3), this product would be $(\phi + 2)nm$. It is, of course, obvious that any multiplication rule of the form

$$n * m = Knm, \tag{9.1}$$

with K a real number, is associative. Now turn to the function g of our Theorem, and assume for simplicity that it is a continuous nonincreasing function such that $g(0) = 1$ and $g(1) = 0$.

Definition. For positive n and m define

$$n * m = n *_g m = N \tag{9.2}$$

where

$$\int_0^1 g(u^{1/N})\,du = \int_0^1 g(u^{1/n})\,du \int_0^1 g(u^{1/m})\,du. \tag{9.3}$$

The left side of (9.3) is strictly decreasing in N, and tends to 0 as $N \to \infty$. By our Theorem, we also know that the left side of (9.3) exceeds the right side when $N = n + m$. Hence the function $N = N(n, m)$ is well defined, and

$$N \geq n + m. \tag{9.4}$$

The multiplication $*$ is associative, since any multiplication rule of the form

$$n * m = f^{-1}(f(n)f(m)) \tag{9.5}$$

is associative. To get a feeling for (9.2), let's consider some particular examples.

If $g(x) = 1 - x^c$ then (9.3) becomes

$$\frac{c}{N+c} = \frac{c}{n+c} \cdot \frac{c}{m+c} \tag{9.6}$$

and

$$N = n + m + \frac{nm}{c}. \tag{9.7}$$

As $c \to \infty$ we recover the usual (associative) law of addition while for $c \to 0$ the $n + m$ becomes negligible and we have essentially (9.1). In fact the multiplication rule (9.7) is quite familiar to algebraists who work with rings [**Mc**, p. 110]. Now examine

$$g(x) = (1 - x)^k, \quad k \text{ an integer}. \tag{9.8}$$

By (8.2), (8.3), and a simple change of variable we have

$$\int_0^1 g(u^{1/n})\, du = \frac{k!}{(n+1)\cdots(n+k)}. \tag{9.9}$$

The corresponding multiplication rule may be described so that it resembles (9.6), to which it reduces when $c = k = 1$: the geometric mean of $N + 1, N + 2, \ldots, N + k$ shall equal the geometric mean of $(n + 1)(m + 1), (n + 2)(m + 2), \ldots, (n + k)(m + k)$.

It is known that in some sense (9.5) describes all possible sensible associative multiplications of the real numbers. See [**Ac**] and the references therein for the details. We conclude by posing a presumably new open question in the realm of nonlinear integral equations. If a given function of two variables

$$N = N(n, m) = n * m \tag{9.10}$$

yields a multiplication that is commutative, associative, and "smooth", and $N \geq n + m$, when and how can (9.3) be solved for g?

REFERENCES

[Ac] J. Aczél, The state of the second part of Hilbert's fifth problem, *Bull. Amer. Math. Soc.* 20 (1989), 153–163.

[Ar] P. Arnoux, Some remarks about Fibonacci multiplication, *Appl. Math. Lett.* 2 (1989), 319–320.

[B] C. Berge, The Theory of Graphs and Its Applications (Chapter 6), Dunod, Paris, 1958.

[B-B] E. F. Beckenbach and R. Bellman, Inequalities, Springer-Verlag, Berlin, 1983.

[B-C-G] E. R. Berlekamp, J. H. Conway and R. K. Guy, Winning Ways, vol. I, Academic Press, London, 1982.

[Ba-C] W. W. Rouse Ball and H. S. M. Coxeter, Mathematical Recreations and Essays, Dover, New York, 1987.

[C] J. H. Conway, On Numbers and Games, Academic Press, London, 1976.

[Ca] B. C. Carlson, Special Functions of Applied Mathematics, Academic Press, London, 1982.

[Cas] J. W. S. Cassels, An Introduction to Diophantine Approximation, Cambridge Univ. Press, Cambridge, 1957.

[Co] I. G. Connell, A generalization of Wythoff's game, *Canad. Math. Bull.* 2 (1959), 181–190.

[Cox] H. S. M. Coxeter, The golden section, phyllotaxis, and Wythoff's game, *Scripta Math.* 19 (1953), 135–143.

[F] A. S. Fraenkel, How to beat your Wythoff's game opponent on three fronts, *Amer. Math. Monthly* 89 (1982), 353–361.

[F-B] _____ and I. Borosh, A generalization of Wythoff's game, *J. Comb. Theory* A 15 (1973), 175–191.

[G] M. Gardner, Penrose Tiles to Trapdoor Ciphers, W. H. Freeman and Company, New York, 1989.

[H-L-P] G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, Cambridge, 1964.

[K] D. E. Knuth, Fibonacci multiplication, *Appl. Math. Lett.* 1 (1988), 57–60.
[K-N] L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences, John Wiley and Sons, New York, 1974.
[Mc] N. H. McCoy, The Theory of Rings, The MacMillan Company, New York, 1964.
[P-S1] H. Porta and K. Stolarsky, The edge of a golden semigroup, Proc. 1987 János Bolyai Math. Soc. Conference on Number Theory 51 (1987), 465–471.
[P-S2] _____, Wythoff pairs as semigroup invariants, Advances in Math. 85 (1991), 69–82.
[P-S3] _____, Some arithmetical semigroups (with A. S. Fraenkel), *Progress in Mathematics* 85 (1990), 255–264.
[S1] K. Stolarsky, Beatty sequences, continued fractions, and certain shift operators, *Canad. Math. Bull.* 19 (1976), 473–482.
[S2] _____, Convection, fundamental inequalities, and new inequalities, *Utilitas Math.* 32 (1987), 181–192.
[W-W] E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., Cambridge Univ. Press, Cambridge, 1962.
[Wy] W. A. Wythoff, A modification of the game of Nim, *Nieuw Archief voor Wiskunde* (2)7 (1905–07), 199–202.

A Topological Approach to Digital Topology

T. Yung Kong* and Ralph Kopperman†, *City College, CUNY, New York, NY 10031*

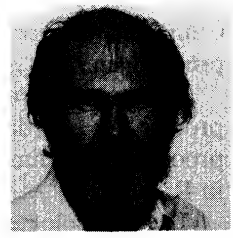
Paul R. Meyer‡, *Lehman College, CUNY*



Yung Kong is Associate Professor of Computer Science at Queens College CUNY. He previously held faculty appointments in computer science at Ohio University and in mathematics at City College, CUNY. His main research interest is in topological and geometrical problems of computer graphics and image processing. He received his doctorate in 1986 from the University of Oxford, England, following three years as a graduate student at the Oxford University Computing Laboratory; his thesis, "Digital Topology with Applications to Image Processing," was directed by Dr. A. W. Roscoe and Professor C. A. R. Hoare. His B.A. is in mathematics (with double first class honors) and is from the University of Cambridge, England.



Ralph Kopperman received his Ph.D. from M.I.T. in 1965 and has taught at CCNY since 1967. His article, "All topologies come from generalized metrics," which appeared in the *Monthly* in Feb. 1988, is also related to the seminar.



Paul Meyer received a B.A. in industrial engineering from Dartmouth College in 1951 and a Ph.D. in mathematics from Columbia University in 1964. Except for leaves in Austin, London, and Padua, he has been at Lehman College since then. He is enjoying doing digital topology after taking his turn as department chair.

The authors are members of the CCNY Seminar on General Topology and Topological Algebra. They met when Kong spoke there in May 1988. That fall he joined the seminar, which the other authors helped begin in 1981. (Another founder was Efim Khalimsky, who started research in digital topological spaces in the late 1960s. This collaboration has led to a better understanding of the relationship between the different approaches to the subject.)

*Present adress: Dept. of Computer Science, Queens College CUNY, Flushing, NY 11367.

†This author's research was supported by a grant from the PSC-CUNY research award program.

‡This author is pleased to thank the Indian Institute of Technology Kanpur and Professor U. B. Tewari for their kind hospitality during the preparation of this paper. We are grateful to Dr. Shobha Madan and Dr. R. K. S. Rathore for their helpful comments.

1. Introduction. Digital topology is a term that has arisen for the study of geometric and topological properties of digital images. In our increasingly digital technology, such questions arise in a wide range of diverse applications, including computer graphics, computer tomography, pattern analysis and robotic design, to mention just a few of the areas of current interest. Despite the name, the traditional approach to this study has been graph-theoretic rather than topological. Although our approach does not abandon graph-theoretical tools, we seek to convince the reader that it is both possible and natural to proceed topologically. The two theories are coextensive for finite spaces (see §2), but the topological approach leads naturally to constructions which, although possible in graph theory, had not previously been considered (see §6 and 7).

To develop a theory for studying images, digital or otherwise, it is natural to start with the Jordan curve theorem, which states that a simple closed curve separates the real plane $\mathbf{R} \times \mathbf{R}$ into exactly two connected components. We want an analog for the digital plane $\mathbf{Z} \times \mathbf{Z}$. The first approach to this, by Azriel Rosenfeld in the late 1960's (see [**22**, **23**] and references there), used the graph-theoretic notion of adjacency to define connectedness. Before giving the definitions necessary for a precise description of Rosenfeld's theorem (for which, by the way, we give a topological proof in §8), let us outline the topological approach.

In the topological approach to this (§3), one defines a connected topology on $\mathbf{Z}$ regarded as an ordered set. (The usual interval topology on $\mathbf{Z}$ is, of course, discrete and not connected.) This induces a product topology on $\mathbf{Z} \times \mathbf{Z}$ (and on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$, when we come to consider the 3 dimensional theorem). In the usual (real) theory one defines a path (resp. arc) as a continuous function (homeomorphism) on a closed real interval; analogously, define *digital path* and *digital arc* by replacing the real interval by an interval subset of $\mathbf{Z}$ (i.e., a set of the form $\{n, n+1, \ldots, n+k\}$ with this new topology). Our Jordan curve theorem in $\mathbf{Z} \times \mathbf{Z}$ (proved in §5, after the necessary background has been developed) is then an exact analog of the usual theorem for $\mathbf{R} \times \mathbf{R}$, where "connected" means the usual definition of connected with respect to the product topology. It is due to Khalimsky [**4**, **5**]; the proof given here is from [**7**].

This approach also yields a Jordan surface theorem in $\mathbf{Z}^3$ (described in §6). Our proof of the Jordan curve theorem is purely digital (i.e., self-contained), but the proof of the 3 dimensional theorem invokes the classical Jordan-Brouwer theorem in $\mathbf{R}^3$.

These topologies are not homogeneous,[1] but that turns out to be an advantage, as we will see in §7, where the following questions about boundaries in digital pictures are considered:

1. How do more complicated curves separate the digital plane into connected components?
2. Conversely, if we are given only a partition of the plane into connected sets, when are the boundary sets curves? And, if they are curves, how can we recover them?
3. How can we represent this on a computer screen?

[1]A topology on X is called *homogeneous* if for each $x, y \in X$ there is a homeomorphism that maps x to y.

We will wait until §7 to answer these questions, after the relevant definitions have been introduced. Let us just mention here that the answer to the last question also settles the non-homogeneity problem: we display a homogeneous subspace which —like visual reality—does not include the bounding curves between adjacent regions.

Returning now to Rosenfeld's graph-theoretic approach, we need some definitions in order to state his theorem. The *4-neighbors* of $(x, y) \in \mathbf{Z} \times \mathbf{Z}$ are the four points $(x \pm 1, y)$ and $(x, y \pm 1)$; the *8-neighbors* are the eight points consisting of the 4-neighbors and the points $(x \pm 1, y \pm 1)$. If $k = 4$ or 8, a *k-path* is a finite sequence $p_0, \ldots, p_n$ in $\mathbf{Z} \times \mathbf{Z}$ such that p_{i-1} is a k-neighbor of p_i, $1 \leq i \leq n$. A set S is *k-connected* if any two of its points can be joined by a k-path in S. A *Jordan k-curve* is a finite k-connected set which contains exactly two k-neighbors for each of its points. As usual, a *component* is a maximal connected set.

The examples in FIGURE 1 show that neither 4-connectedness nor 8-connectedness give a Jordan curve theorem. The first is a Jordan 8-curve whose complement is 8-connected; the second is a Jordan 4-curve which is not a Jordan 8-curve and whose complement has three 4-connected components. However, by combining these two notions of connectedness, Rosenfeld proved (see [22] and references there):

THEOREM 1.1 (Rosenfeld). *A Jordan k-curve with at least five points separates* $\mathbf{Z} \times \mathbf{Z}$ *into exactly two k'-components where* $\{k, k'\} = \{4, 8\}$.

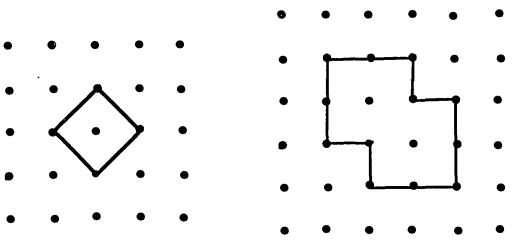


FIG. 1. A Jordan 8-curve and 4-curve.

Note that the definition of connectedness changes when the curve changes.

Since finite topological spaces behave very differently from the usual ones, we begin (§2) by developing techniques for such spaces in some detail. We are primarily interested in finite T_0-spaces, but we give some indication of where the ideas are valid more generally.

This paper is more of an introduction than a survey (for surveys see Rosenfeld [22] and Kong and Rosenfeld [13]). In [15] Kovalevsky proposed an approach similar to ours, but without proofs; see also [16]. Since we are interested in the foundations of the subject and the relationship between the topological and the graph-theoretical approaches, §2 and 3 contain more detail than is needed for the sequel.

2. Finite topological spaces. Recall that a topology on a set X is a collection of subsets of X including X and $\varnothing$ and closed under finite intersections and arbitrary unions. A set is closed if its complement is open, so that closed sets are

preserved by finite unions and arbitrary intersections. For $x \in X$ let $N(x)$ denote the intersection of all open neighborhoods of x, which of course is not a neighborhood in general.

If X is finite then open sets and closed sets behave similarly with respect to intersection and union and each point has a smallest open neighborhood, namely $N(x)$. We are primarily concerned with finite spaces but much of this section is valid more generally for a class of spaces introduced by Alexandroff [**1**], who called them 'discrete': an *Alexandroff topology* is one in which every intersection of open sets is open, or, equivalently, in which each $N(x)$ is open. (It should be noted that some recent authors, e.g., Johnstone [**2**], include T_0 in this definition.[2]) We also consider an intermediate class of spaces between finite and Alexandroff: call a space *locally finite* if each point is in a finite open set and in a finite closed set, but note that this is stronger than the usual definition.

Thus if X is an Alexandroff space we have for $x \in X$, $A \subset X$:

$$x \in \text{cl } A \text{ implies } x \in \text{cl}\{y\} \text{ for some } y \in A. \tag{1}$$

It follows from this that an Alexandroff topology is completely determined if one knows each $N(x)$ or, alternatively, each $\text{cl}\{x\}$. What is more interesting is that, by (1), such a topology is determined by knowing the subspace topology on each two point subset. (Actually, (1) is equivalent to the Alexandroff property: a topology satisfies (1) iff it is Alexandroff.) Note that for x, y in any topological space we have:

$$x \in \text{cl}\{y\} \text{ iff } y \in N(x) \tag{2}$$

but this is not enough to determine the topology in spaces where $N(x)$ is not necessarily open.

So let us look (FIGURE 2) at the possible topologies on a two point set $\{x, y\}$ (by drawing $N(x)$ and $N(y)$ in each case). These are the discrete, indiscrete, and two copies of a topology called the *Sierpinski topology*. (The diagram is drawn this way to be consistent with the specialization order to be introduced shortly.) Thus if X is any T_0-space with $x, y \in X$ we have:

$$\{x, y\} \text{ is connected iff the subspace topology on } \{x, y\} \text{ is Sierpinski.} \tag{3}$$

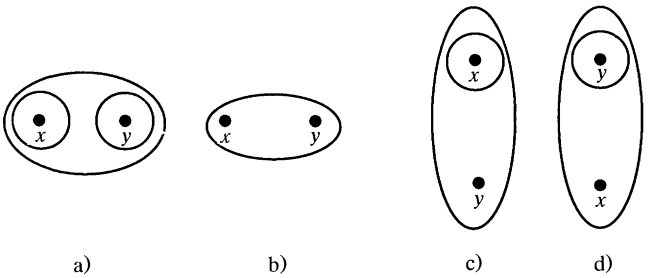


FIG. 2. Topologies on a two point space.

[2]Recall that a topological space X is called a *T_0-space* (resp., *T_1-space*) *if, whenever x and y are distinct points in X, $x \notin cl\{y\}$ or (resp. and) $y \notin cl\{x\}$.*

There is a correspondence between topological spaces and directed graphs, which can be obtained from the *specialization order* (introduced by Alexandroff [1] and discussed, for example, in [2]): For any topological space, $x \leq y$ iff $x \in \mathrm{cl}\{y\}$ defines a reflexive transitive relation. This relation is a partial order (i.e., anti-symmetric) iff the space the space is T_0 (and then it is called the specialization order). Thus in any topological space, $\mathrm{cl}\{y\} = \{x: x \leq y\}$ and, by (2), $N(x) = \{y: x \leq y\}$. Also, the closed points (resp. open points) are precisely the minimal (maximal) elements. If a T_0-space is finite, or more generally locally finite, then each point lies below a maximal element. Stated topologically this says that in such spaces the set of open points is dense in the following strong sense; we call it *pointwise dense*:

$$\cup\{\mathrm{cl}\{x\}: \{x\} \text{ is open}\} = X.$$

It is useful to represent the specialization order on X by making X a digraph (= directed graph) by putting an edge directed, say, from x to y whenever $x \leq y$. In this way each T_0-topology determines a unique transitive digraph. Conversely, each partial ordering determines a unique Alexandroff topology. However, more general topologies in which $N(x)$ need not be open are not determined by their specialization order. For example, on an infinite set, the discrete topology, the cofinite topology, and any topology between these two, all have the same specialization order, namely equality.

For finite X, however, this gives a one to one correspondence between T_0-topologies on X and partial orderings on X. If we drop the T_0 axiom, it gives a one to one correspondence between topologies on X and transitive digraphs on X (where here it is understood that both $x \to y$ and $y \to x$ are possible).

We also use the *connectedness graph* of a topology, an undirected graph in which an edge joins x and y iff $\{x, y\}$ is topologically connected. This can be obtained from the digraph above by just forgetting the direction of the edges. In this context define the *adjacency set* $\mathscr{A}(x)$ of a point x as $\{y \neq x: \{x, y\}$ is connected$\}$. In any topological space (don't need T_0 here):

$$\mathscr{A}(x) = (N(x) \cup \mathrm{cl}\{x\}) - \{x\}. \tag{4}$$

These concepts are interesting only in non-T_1-spaces because: X is T_1 iff $\mathscr{A}(x) = \varnothing$ for each x iff $N(x) = \{x\}$ for each x iff the specialization order relation is equality.

We conclude this section with two results on continuity of functions. We omit the easy proofs; see [19].

PROPOSITION 2.1. *A function defined on a T_0-Alexandroff space is continuous iff it preserves specialization order.*

PROPOSITION 2.2. *If f is defined on $X \times Y$ and Y is locally finite, then f is jointly continuous iff it is separately continuous.*

3. The digital line and digital arc connectedness. The key to the topological construction of the digital plane is getting the right topology on $\mathbf{Z}$. This definition was published in Russia by E. Khalimsky in 1970 [5] but not widely circulated outside Russia until recently. FIGURE 3 shows the topology (by showing each $N(x)$); observe that the points alternate being open and closed. (A more formal description is given in [7, §10].) It can be defined on $\mathbf{Z}$ or on an interval subset of $\mathbf{Z}$. We refer to this topology, and the associated product topologies on $\mathbf{Z}^n$, as

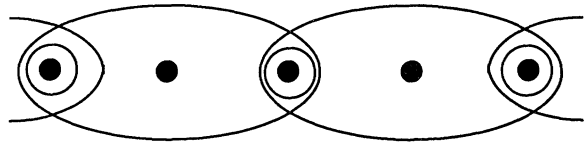


FIG. 3. A portion of a digital line, showing the minimal neighborhoods of each point.

Khalimsky topologies; Marcus [18] independently constructed a topology which agrees with a Khalimsky topology on **Z**, but differs on $\mathbf{Z}^2$.

There is a sense in which this topology is uniquely determined. Like the real line **R**, this space satisfies the following condition, which justifies regarding it as a digital line.

DEFINITION 3.1. A connected ordered topological space (*COTS*) *is a connected topological space X with this property*: *for any* x_1, x_2, x_3 *distinct points in X, there is an i such that* x_j *and* x_k *lie in different components of* $X - \{x_i\}$, *where* $\{i, j, k\} = \{1, 2, 3\}$. (*There is no separation*[3] *assumption made on X*).

It can be shown [7, Theorem 7] that each COTS induces a total order on the underlying set such that: if x is not an end point, the components of $X - \{x\}$ are $\{y: y > x\}$ and $\{y: y < x\}$ (call these sets $U(x)$ and $L(x)$ resp.)[4] This COTS order is unique up to inversion. Note that it differs from the specialization order;[5] in the case of the real and digital lines the COTS order is the usual one.

The next proposition shows that the topology given by FIGURE 3 is unique in the sense that any locally finite COTS must have this form; in the finite case there is the question whether the first point is open or closed. In fact, every locally finite COTS (more generally, every Alexandroff COTS) with at least three points is homeomorphic to an interval subset of **Z**.

PROPOSITION 3.2 [7, Prop. 8]. *In a COTS with at least three points*,[6] *each point is either open or closed. A point x is closed* (*resp. open*) *iff* $U(x)$ *and* $L(x)$ *are open* (*closed*). *If two points are order-adjacent then one of them is open and the other closed. If x has a successor, but no immediate successor, then x is closed.*

Proof. If a COTS has at least three points, then it has a point which is not an end point; we show that any such x is open or closed (a separate argument is required for end points). If A, B separate $X - \{x\}$ then

$$\operatorname{cl} A = A \quad \text{or} \quad \operatorname{cl} A = A \cup \{x\}$$

and

$$\operatorname{cl} B = B \quad \text{or} \quad \operatorname{cl} B = B \cup \{x\}.$$

[3]We remind the reader that the term *separation* has two different standard meanings: separation axioms (T_0, T_1, etc.) and separation of a set which is not connected (see, for example, [3]).

[4]Conversely, any connected topological space with a total ordering which is related to the topology as in the previous sentence is a COTS.

[5]From here on the usual order notations will be used for the COTS ordering. There should be no confusion with the specialization ordering because the latter will appear only in the digraph and connectedness graph which were constructed from it.

[6]This hypothesis is necessary because the two point indiscrete space is a COTS.

Both A and B must behave similarly (otherwise we have a separation of X). If A, B are both closed then $\{x\}$ is open; otherwise $\text{cl } A \cap \text{cl } B = \{x\}$ is closed, and A, B are open.

If x, y are order-adjacent, say $x = y^+$, the successor of y, we show one is open and the other closed. Say x is closed, so that $U(x)$ is open. Then $\text{cl } U(x) = U(x) \cup \{x\} = U(y)$ is closed, and $\{y\}$ is open; here again a modification is necessary if x is an end point. A similar argument works if we assume one of the points open.

For the last sentence, show $U(x)$ is open by noting that, for each $y > x$, $U(y)$ or $U(y) \cup \{y\}$ is open, and $U(x)$ is a union of these.

In the classical theory a path (resp. arc) in a topological space is a continuous function (homeomorphism) from a closed real interval. We define a *digital path* (resp. *digital arc*) in a topological space to be the range of a continuous function (homeomorphism) from a finite COTS, i.e., from a finite interval in $\mathbf{Z}$. If it is clear from the context, we sometimes omit the adjective digital.

The following properties will be useful in Jordan curve theory.

THEOREM 3.3 [7, Theorem 13]. *A finite topological space is connected iff it is digitally arc-connected iff it is digitally path-connected.*

If C is a finite connected set containing distinct points x and y, then C is a digital arc with end points x and y iff $|\mathscr{A}(x)| = |\mathscr{A}(y)| = 1$ and $|\mathscr{A}(w)| = 2$ for any $w \in C - \{x, y\}$ (i.e., iff there is no "extra" connectedness).

Proof. The first statement follows immediately from the following characterization of COTS-arc [7, Proposition 13(b)]: In any topological space a set is minimal among connected sets containing points x and y iff it is a COTS-arc with end points x, y. The second follows because, in a COTS, $\mathscr{A}(w)$ is the set of points order-adjacent to w. The theorem is actually valid more generally for Alexandroff spaces.

A more complete discussion of COTS with detailed proofs can be found in [7, Sections 2, 3]. Although it is not needed for the sequel, let us note how the last sentence of Proposition 3.2 relates this theory to a now classical theory of connected ordered T_1-spaces (see [9] for a recent survey): considering only T_1-spaces is equivalent to considering only orderings without order-adjacent points (see also [7, Proposition 9]).

4. The topology of the digital plane. A space $X \times Y$ with the product topology, where X and Y are locally finite COTS (i.e., interval subsets of $\mathbf{Z}$) with at least three points, is called a *digital plane*. Point (x, y) is called *pure* if $\{x\}$ and $\{y\}$ are either both open or both closed, *mixed* otherwise (i.e., one open and the other closed). Define $\textit{border}(X \times Y) = \{(x, y)$: either x or y is an end point$\}$. The border of a finite plane will play an important role in the proof of the digital Jordan curve theorem. On the other hand, it is sometimes convenient to have a plane with no border, in which case we use $\mathbf{Z} \times \mathbf{Z}$.

Let us begin by considering the connectedness graph of a digital plane, a portion of which is shown in FIGURE 4. (We omit the details of the construction, but see §6 where they are given for the analogous construction in $\mathbf{Z}^3$.) Several properties will be useful.

1. The two different kinds of pure points behave similarly with respect to connectedness. The adjacency set of a pure (resp. mixed) point which is not a border point has eight (four) points.

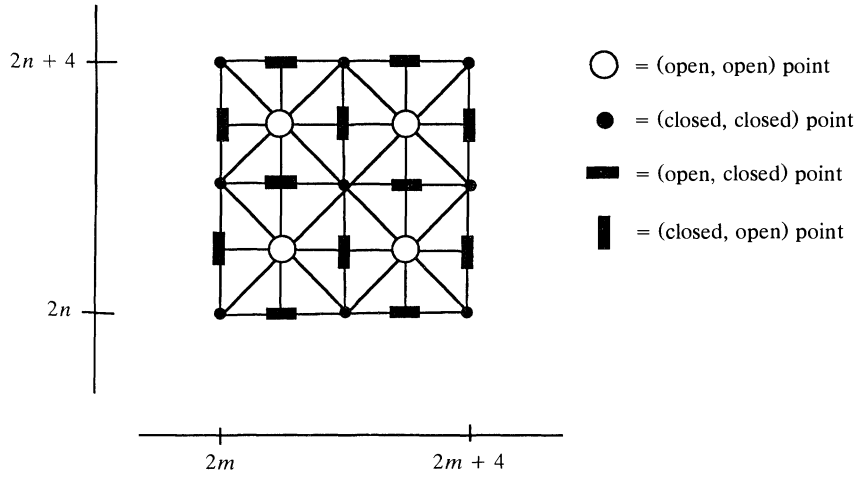


FIG. 4. Connectedness graph of a portion of a digital plane.

2. In constructing a digital path one can follow any sequence of points as long as adjacent points on the path are connected in the graph. Thus, by the preceding theorem, the connectedness graph is sufficient to determine which sets are connected.

3. For an arc, however, there can be no "extra" connectedness (as noted in the preceding theorem); thus an arc cannot turn at a mixed point (since the pure point before the mixed point would be connected to the pure point after the mixed point).

Although it will not be needed in the sequel, one can construct the digraph of the specialization order from the (undirected) connectedness graph by directing the edges as follows: edges go from closed points and towards open points. Since every edge contains a pure point, this suffices to direct all of the edges. Note that mixed points are saddle points in this ordering.[7]

Thus far we have considered digital lines as objects arising naturally from the COTS property and digital planes as products of such lines. It is important to note that there is an alternative construction of digital spaces as suitable quotients of real spaces. This will play an essential role in §6 when we use the Jordan-Brouwer theorem for real 3-space in a digital proof and in §7 when we consider how to display digital images. Define f: $\mathbf{R} \to \mathbf{Z}$ by $f(2n) = 2n$ and $f(x) = 2n + 1$ whenever $2n < x < 2n + 2$. In the quotient topology on $\mathbf{Z}$ odd points are open and even points are closed, so the quotient topology is the COTS topology. Now define g: $\mathbf{R}^2 \to \mathbf{Z}^2$ by $g = f \times f$. Again, the quotient topology coincides with the digital topology.

Alternative diagrams of this quotient topology on digital lines and planes are obtained by looking anew at the diagram of a partitioning of a line (FIGURE 5) and seeing each dot as a closed point and each subinterval between two dots as an

[7]For an Alexandroff topology, reversing the specialization order gives a new Alexandroff topology in which the new open sets are the old closed sets. These topologies are not homeomorphic in general.

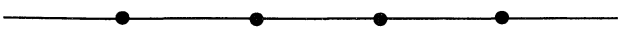


FIG. 5. The digital line as the quotient of the real line.

open point. Thus a set is open in **Z** if it "looks open"; i.e., the corresponding set is open in **R**. The corresponding diagram of the digital plane looks like FIGURE 6. Here the open points are represented by the interiors of the rectangles, the closed points by their corners, and the mixed points by their open sides. One's intuition about these digital structures can be improved by using FIGURE 6 to derive the connectedness graph given in FIGURE 4.

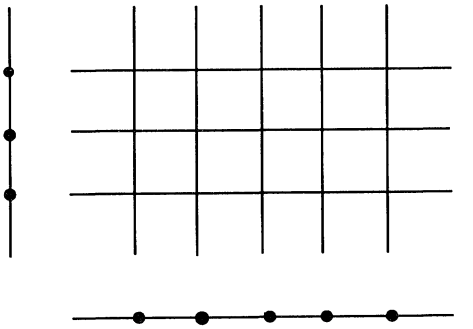


FIG. 6. The digital plane—quotient of the real plane or product of digital lines.

This gives further insight into why a digital arc should not turn at a mixed point: to exit from the interior of a square, an arc goes through an edge *or* a corner, but not an edge and *then* a corner.

5. The digital Jordan curve theorem. We say that a *digital Jordan curve* is a finite connected set J with $|J| \geq 4$ such that $J - \{j\}$ is a digital arc for each $j \in J$. By [7, Prop. 17(a)] this is equivalent to saying that $J \neq \emptyset$ is connected with $\mathscr{A}(j) \cap J$ a two point discrete set for each $j \in J$. For use in the next section we note that the second characterization extends to three dimensional space as well. We now sketch the proof of the digital Jordan curve theorem; for a more formal treatment with all of the details see [7, Section 4]. Let $X = X_1 \times X_2$ be a finite digital plane. We will use the fact that, for any non-border point j, $\mathscr{A}(j)$ is a digital Jordan curve and $\mathscr{A}(j) - J$ has exactly two components. Since this proof utilizes the border, we begin by proving

THEOREM 5.1. *If C is an arc in X, then $X - C$ and* border$(X) - C$ *have the same number of components and these correspond by set inclusion.*

Proof. Let $\mathscr{B} = border(X)$ and consider the map Ψ from the set of components of $\mathscr{B} - C$ to the set of components of $X - C$ defined by letting $\Psi(W)$ be the component in $X - C$ which contains W. We show that Ψ is a bijection by showing:

(i) Each component of $X - C$ meets $\mathscr{B} - C$.

(ii) If D is an arc in X that meets more than one component of $\mathscr{B} - C$, then D meets C.

To prove (i) by induction on $|C|$, assume that we have a counterexample in which C is as short as possible and suppose that the component of some point p does not meet $\mathscr{B}$. If f is an endpoint of C, then $C' = C - \{f\}$ is a shorter arc, so the component of p in $X - C'$ meets $\mathscr{B}$. There is an arc D in $X - C'$ connecting p to some $b \in \mathscr{B}$, and f must be on D. We get a contradiction by constructing an arc D' in $X - C$ also connecting p to b. We construct D' from $D - \{f\}$ by using a portion of $\mathscr{A}(f) - C$ to "bridge the gap" which arose when f was removed from D. FIGURE 7 shows one possible configuration where f is a mixed point. To justify using $\mathscr{A}(f) - C$ in this fashion, note that $|\mathscr{A}(f) \cap C| = 1$ and that $\mathscr{A}(f)$ is a Jordan curve, so that $\mathscr{A}(f) - C$ is an arc by [7, 17(a)]. The remaining cases and other details can be found in [7, Lemma 18].

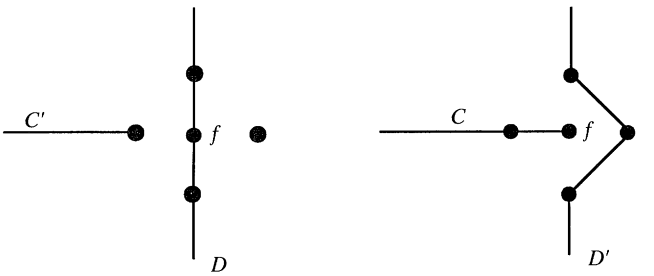


FIG. 7. Construction of D' for the case in which f is a mixed point not on $\mathscr{B}$.

If (ii) fails, let X_1, X_2, C, D be a minimal counterexample; note that minimality implies that C and D meet the border at precisely their endpoints. If $|X_1| = |X_2| = 3$ the result is easy to verify, so we may assume, say, $|X_2| > 3$. Let y be the initial point of X_2 and consider a second example obtained by deleting the bottom row of the original plane; i.e., replace X_2 by $X_2 - \{y\}$. By the minimality, this is no longer a counterexample. This yields a contradiction, the details of which are in [7, Lemma 19].

THEOREM 5.2. *If J is a digital Jordan curve in X which does not meet* border(X), *then $Z^2 - J$ has exactly two components.*

Proof. To show that $X - J$ has at most two components, fix $j \in J$ and note that $J - \{j\}$ is an arc which does not meet *border*(X). By the preceding theorem, $X - (J - \{j\})$ is connected; thus any point in $X - J$ can be joined to j by an arc in $X - J$. Such an arc must meet $\mathscr{A}(j) - J$. This shows that any point in $X - J$ can be joined to $\mathscr{A}(j) - \{j\}$ by an arc in $X - J$. Since $\mathscr{A}(j) - J$ has two components, it follows that $X - J$ has at most two components.

To show that $X - J$ has at least two components, we must find an "inside" point (i.e., a point which cannot be connected to *border*(X) by an arc in $X - J$). To utilize the previous theorem, we shorten X, as we did in the proof of (ii), by "moving up" the bottom edge until it touches J. By modifying J, we then construct an arc C which separates the border of this shortened plane X'. By the previous theorem $X' - C$ has at least two components, and the existence of the desired inside point can be deduced from this. The many details missing from this sketch can be found in the proof of [7, Theorem 21].

Although this theorem was formulated in a finite plane in order to use the border in the proof, as an obvious corollary we get an analogous theorem for (finite) Jordan curves in $\mathbf{Z}^2$.

THEOREM 5.2. *If J is a digital Jordan curve in $\mathbf{Z}^2$, then $\mathbf{Z}^2 - J$ has exactly two components.*

6. The Jordan surface theorem. There are also three dimensional questions in image processing; the third dimension can be spatial of course (e.g., X-ray tomography), but it can also be time (e.g., a "movie" is a sequence of changing two dimensional images) or a "gray scale function", which assigns to each pixel in a two dimensional digital image a numerical value measuring the shade of gray assigned to that pixel. Since the topology for the digital plane was a product topology, there is such a topology for $\mathbf{Z}^3$. We use this to prove a digital analog of the classical Jordan-Brouwer theorem which asserts that every compact 2-manifold separates $\mathbf{R}^3$ into two connected components.

Let $X = X_1 \times X_2 \times X_3$, where each X_i is a finite COTS; let τ denote the product of the COTS topologies. A *digital m-simplex* in X is a set of $m + 1$ mutually connected points. It is convenient to use the terms *dyad* (resp. *triad*, *tetrad*) for the cases $m = 1$ (resp., 2, 3).

As in the digital plane (§4), there are just two types of adjacency sets, for pure points and for mixed points (definitions analogous to those given earlier). This time we verify this by constructing $\mathscr{A}(p)$ for the various types of non-border points $p = (x, y, z)$:

1. If p is closed then[8]
$$N(p) = \{x^-, x, x^+\} \times \{y^-, y, y^+\} \times \{z^-, z, z^+\}$$
and $\mathrm{cl}\{p\} = \{p\}$, so that $|N(p) \cup \mathrm{cl}\{p\}| = 27$ and $\mathscr{A}(p)$ has 26 points.
2. If p is open, then the roles of N and cl are reversed, but the result is the same.
3. If p is mixed, say (open, open, closed), then
$$N(p) = \{x\} \times \{y\} \times \{z^-, z, z^+\}$$
and
$$\mathrm{cl}\{p\} = \{x^-, x, x^+\} \times \{y^-, y, y^+\} \times \{z\}$$
so that $\mathscr{A}(p)$ has 10 $(= 2 + 8)$ points.

The connectedness graph can again be used to determine which sets are connected. In FIGURE 8 we illustrate the connectedness relationships in a unit cube of X; with an appropriate rotation any unit cube of X can be reduced to this configuration.

Define a *digital Jordan surface* in X to be a τ-connected subset S of X which does not meet the border and such that $\mathscr{A}(p) \cap S$ is a digital Jordan curve for each $p \in S$. The next lemma [**14**] gives some properties which begin to justify calling such sets surfaces.

[8]Notation: x^- (resp. x^+) denotes the predecessor (successor) of x in the COTS-ordering.

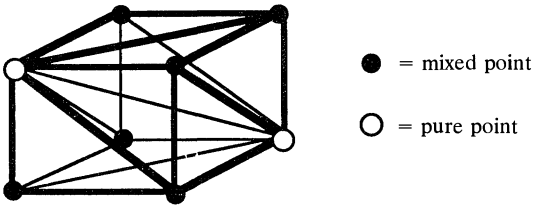


FIG. 8. The connectedness graph of a unit cube in digital three space.

LEMMA 6.1. *If S is digital Jordan surface, then each dyad in S is contained in exactly two triads in S, and S contains no tetrads.*

We sketch the proof of the separation theorem for three dimensional space; full details can be found in [**14**]. The proof utilizes the construction of a *polyhedral analog* in $\mathbf{R}^3$ of the digital Jordan surface S. This construction has been generalized by Kong and Khalimsky [**10**] to arbitrary locally finite T_0-spaces; we use their notation here. This separation theorem has been generalized to n-dimensional digital space by T. Yung Kong (personal communication). The concept of approximating polyhedra by such finite spaces goes back to Alexandroff [**1**].

THEOREM 6.2. *If S is a digital Jordan surface in X then $X - S$ has exactly two τ-components.*

Proof. If we regard X as a subset of the integer lattice points of $\mathbf{R}^3$, then each digital simplex in X determines a real simplex in $\mathbf{R}^3$ (take the euclidean convex hull). Now each subset S of X is a union of digital simplexes, and in fact determines a simplicial complex $K(S)$ in $\mathbf{R}^3$. We are interested in $|K(S)|$, the associated polyhedron. Let υ denote the usual euclidean topology for $\mathbf{R}^3$. One can show that $|K(S)|$ is a compact 2-manifold in $\mathbf{R}^3$. By the classical Jordan-Brouwer theorem, $\mathbf{R}^3 - |K(S)|$ has two υ-components, I_e and E_e. Put $I = X \cap I_e$ and $E = X \cap E_e$.

It remains to show that I and E are the τ-components of $X - S$. That I and E are τ-separated follows readily from the fact that I_e and E_e are υ-separated. We show that I is non-void and τ-connected by constructing a continuous surjection f: $(I_e, \upsilon) \to (I, \tau)$. (A similar argument works for E). We construct f as a composition of ρ: $(I_e, \upsilon) \to (|K(I)|, \upsilon)$ and π: $(|K(I)|, \upsilon) \to (I, \tau)$. In §4 we constructed a quotient mapping $\mathbf{R}^2 \to \mathbf{Z}^2$. The desired π is the restriction to $|K(I)|$ of the analogous quotient mapping $\mathbf{R}^3 \to \mathbf{Z}^3$. Since some points in I are τ-connected to points in S, we must retract I_e onto $|K(I)|$ before taking the quotient; see [**14**] for the details of this construction of ρ.

For two dimensions our topological theory is equivalent to the graph-theoretic version in the following sense. Our Jordan curves (and Jordan surfaces) yield graphs which are "normal digital pictures" in the sense of Kong and Roscoe [**12**], and their approach gives a graph-theoretic proof for our Jordan curve theorem. Conversely, there is a topological proof of the Rosenfeld Jordan curve theorem (see §8 for details). In three dimensions, however, the topology yields a new Jordan surface theorem. Originally, 6, 18 and 26-adjacency (definitions analogous to those for 4 and 8-adjacency in two dimensions) were studied and Jordan surface theorems were developed [**21**, **20**, **11**] for (6, 26), (26, 6), (6, 18), and (18, 6) (see [**13**, §6] for further references).

7. Boundaries in digital planes. Let us look more carefully at the question of how pixels are represented mathematically. In a black and white image, a pixel takes one of two values, so pixels are often represented as points in $\mathbf{Z}^2$ and the image as a $\{0, 1\}$-valued function on the set of pixels. But an actual pixel has area. How does one deal with points that have area? Some models regard all of the elements as points in $\mathbf{Z}^2$, while others regard a point in $\mathbf{Z}^2$ as representing the unit square in $\mathbf{R}^2$ of which it is the center.

We now have a model in which some elements of $\mathbf{Z}^2$ act like points (the closed points), some like segments (the mixed points), and some act like areas (the open points), but only the latter are actually displayed (i.e., correspond to pixels). As noted in §4, our topology arises as the quotient topology from a quotient mapping of the real plane onto the digital plane, in which a closed unit square in $\mathbf{R}^2$ is partitioned into 9 equivalence classes: the 4 corner points as singletons, the 4 open segments of the border, and, finally, all of the interior points as the ninth class.

In order to consider the questions about boundaries from §1, we need some definitions. A *scene* in a digital plane X is a partition of X into connected subsets. A *separator* in X is either a digital Jordan curve which does not meet $border(X)$ or a digital arc which meets $border(X)$ at precisely its endpoints. We saw in §5 that, if J is a separator, then $X - J$ has exactly two components; furthermore, it is not difficult to see that J is minimal with respect to this property.

Example 8 of [**8**] shows that we need some conditions on the elements of a scene in order that their boundaries be unions of separators. To that end we define a *robust scene* to be a scene, each of whose elements has connected interior and is contained in the closure of its interior.

Defining a *cartoon* to be a union of closed separators, we get a natural correspondence between robust scenes and cartoons. Before looking at this theorem, let us note some important properties of closed separators. By the *boundary* of set A we mean the topological boundary, i.e., cl A − int A, which we denote by *bdry* A.

PROPOSITION 7.1 [**8**, Proposition 10]. *The following are equivalent for a separator* J:

1. J *is the boundary of one of the components of* $X - J$.
2. J *is closed.*
3. J *contains no open points.*
4. J *is nowhere dense.*

The last two properties are actually equivalent for any subset of a locally finite space (recall that the set of open points is pointwise dense in such spaces).

THEOREM 7.2 [**8**, Theorem 13]. *If* C *is a cartoon, then there is a robust scene* Π *such that*

$$C = \bigcup\{bdry\, A\colon A \in \Pi\}.$$

Further, $\{\text{int } A\colon A \in \Pi\}$ *is uniquely determined by* C. *Conversely, if* Π *is any robust scene then*

$$C = X - \bigcup\{\text{int } A\colon A \in \Pi\}$$

is the unique cartoon for which $C = \bigcup\{bdry\, A\colon A \in \Pi\}$.

The fact that a cartoon does not uniquely determine the elements of a scene, but only their interiors, reflects a reality of vision: we do not "see" the bounding

curves between adjacent regions; rather, we infer them from the regions they bound. This ambiguity about what to do with the bounding curves disappears when we consider the third question: how should one display these images? Our answer is to display only the subspace consisting of the topologically open points. Define the *open screen* $\mathscr{P}$ in a digital plane to be the subset consisting of all the open points. In this approach only points of the open screen are actually represented as pixels; other points are not displayed.

Note that this also resolves a difficulty mentioned in §1; although the digital plane is not homogeneous, the open screen is, as the setting for a picture should be. Also, if we consider this from the perspective of the quotient mapping of the real plane onto the digital plane, we see that it is the grid lines in the real plane which correspond to the points not represented by pixels in the open screen. This grid is a set of measure zero.

The trace of a robust scene on the open screen is called its *display*; more formally, if Π is a robust scene, then

$$D_\Pi = \{A \cap P\colon A \in \Pi\}$$

is its display. Our next theorem provides a unified answer to the three questions about boundaries raised in §1.

THEOREM 7.3 [**8**, Theorems 18 and 19],

1. *Displays and cartoons of robust scenes uniquely determine each other.*
2. *Each display of a robust scene is a partition of the open screen* $\mathscr{P}$ *into "4-connected" sets. Conversely, each such partition is the display of some robust scene.*

Since $\mathscr{P}$ is still a subspace of X, the concept "4-connected in $\mathscr{P}$" requires a slightly different definition of 4-neighbor from that given in §1: the 4-neighbors of (x, y) are the points $(x \pm 2, y)$ and $(x, y \pm 2)$.

A robust scene determines a unique cartoon and a unique display; furthermore, a display and a cartoon uniquely determine each other. However, a robust scene is not uniquely determined by either its display or its cartoon, because one must decide where to put the bounding curves. One method of constructing the robust scene in either of these cases: enumerate the connected sets in the partition (i.e., the 4-connected sets (in the case of a display) or the topological components of the complement of the cartoon), say, $Q_1 \ldots Q_n$. Let $A_i = \operatorname{cl} Q_i - \cup\{\operatorname{cl} Q_j\colon j < i\}$. Then in both cases $A_1, \ldots, A_n$ turns out to be an appropriate robust scene.

To summarize then, we have three ways to specify a digital image, and they are equivalent in the sense that the classes determined by each of them are coextensive and each determines the other two essentially uniquely. One can specify the bounding curves (cartoons) or specify the connected regions, either as topologically connected sets in the digital plane (robust scenes) or as 4-connected subsets of the open screen (displays).

This latter characterization is reminiscent of the Rosenfeld theorem stated in §1. We conclude by giving a topological proof of this [**8**, Theorem 20] in the next section, but before getting to that let us describe briefly another more general topological approach to some of these issues. Kronheimer [**17**] defines a *grid* Δ on a topological space S to be a decomposition in which the union of the open elements is dense and the projection $S \to S/\Delta$ is open. (Thus the Khalimsky

topology forms a grid on $\mathbf{R}^n$.) A *trace space* is a space in which the set of all open points is dense; any trace space can be reduced to a unique minimal trace space: roughly, the quotient obtained by identifying as many non-open points as possible, subject to the identification map being open. Call a trace space *digital* if it is both minimal and semi-regular. It follows that any regular Hausdorff space with a dense discrete subset (e.g., any Hausdorff compactification of $\mathbf{N}$) is digital. He then constructs, on an arbitary set E, a *universal* digital space (by topologizing the collection of all filters on E with a variant of the Wallman topology), universal in the sense that every digital space with a trace equipotent to E can be densely embedded in the universal space.

8. A topological proof of the Rosenfeld Theorem. We prove the Rosenfeld theorem stated earlier by embedding a Rosenfeld plane, $R = \mathbf{Z}^2$ with no topology, into a digital plane, $X = \mathbf{Z}^2$ with the Khalimsky topology, via the *slant map* S (so called because the embedding uses a 45° rotation) defined by:

$$S(x, y) = (x + y, -x + y).$$

If U denotes the set of pure points in X then *range* $S \subset U$ because $(x + y) + (y - x) = 2y$ is always even.

Recalling that connectedness in X means topological connectedness, whereas in R it means k-connectedness for $k = 4$ or 8, let us look at images under S of 4 and 8-neighbors of (x, y) in R as shown in FIGURE 9.

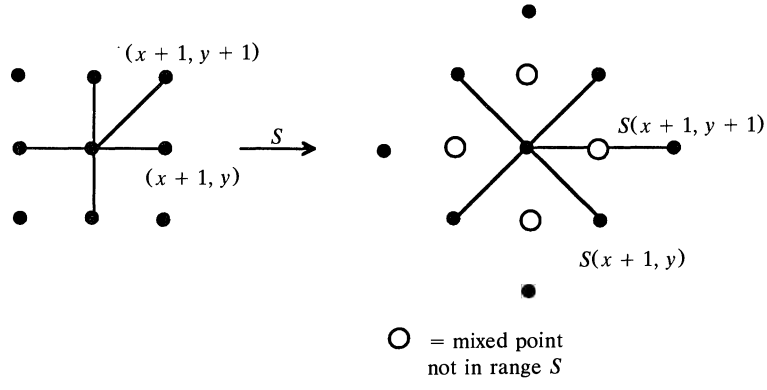


FIG. 9. Images of 4 and 8-neighbors under the slant map S.

Thus: q is a 4-neighbor of p iff $S(q) \in \mathscr{A}(S(p))$ and q is a proper 8-neighbor of p iff there is a unique mixed point on the (necessarily horizontal or vertical) line joining $S(p)$ and $S(q)$.

If J is a 4-path in R, let $J^* = S(J)$. If J is an 8-path, form J^* by adjoining to $S(J)$ the mixed point between $S(p_i)$ and $S(p_{i+1})$ whenever p_{i+1} is a proper 8-neighbor of p_i on J.[9] If J is a k-path (resp. k-arc, Jordan k-curve) in R then J^*

[9] In the case where J is a Jordan 8-curve, one must also do this between the last point and the first point of J.

is a topological path (arc, Jordan curve) in X. If H, J are disjoint in R, with one a 4-path and the other an 8-path, then note that H^* and J^* are disjoint in X.

Now to sketch the proof of the theorem (full details may be found in [**8**, Theorem 20]), let J be a Jordan k-curve in R which has at least five points and does not meet the border. Since $X - J^*$ has exactly two components, I and E, it remains to show that $S^{-1}(I)$ and $S^{-1}(E)$ are the k'-components of $R - J$.

We first show that $S^{-1}(I)$ and $S^{-1}(E)$ are k'-separated. If not, there is a k'-arc H which joins them and is disjoint from J. But then H^* and J^* are disjoint in X, which contradicts the separation of I and E.

To show $S^{-1}(I)$ non-void it suffices to verify that I contains a pure point, which is seen by using the hypothesis $|J| > 4$.

To show that $S^{-1}(I)$ is k'-connected, we let $p, q \in S^{-1}(I)$ and construct a k'-path joining them. Now $S(p), S(q) \in I$, so there is a topological path C in I joining them. Since C does not meet J^*, $S^{-1}(C)$ is an 8-path in $S^{-1}(I)$ joining p and q, and we are done if $k' = 8$. If $k' = 4$, we must "square the corners" of $S^{-1}(C)$ to make it a 4-path. For the 8-neighbors on $S^{-1}(C)$ illustrated in FIGURE 10, we need to know that either u or v is available (i.e., is in $S^{-1}(I)$), so that it can be adjoined to the path. Since $S^{-1}(I)$ and $S^{-1}(E)$ are k'-separated, $u, v \notin S^{-1}(E)$.

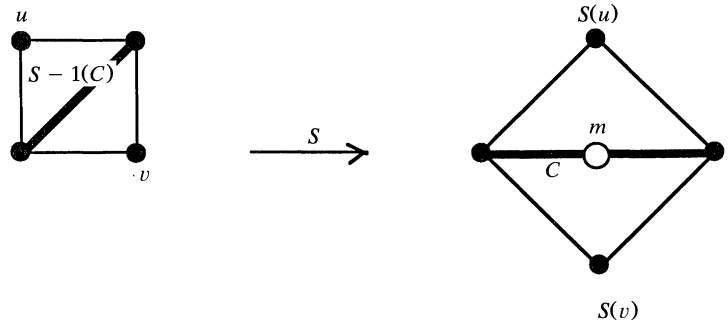


FIG. 10. To convert $S^{-1}(C)$ from an 8-path to a 4-path we 'square the corner' by adjoining either u or v.

If $u, v \in J$, then $S(u), S(v) \in J^*$ and the mixed point m must be on J^*, which is not possible since C and J^* are disjoint. Thus either u or v can be adjoined to $S^{-1}(C)$ to "square the corner". This can be done for each pair of proper 8-neighbors. This shows that $S^{-1}(I)$ is k'-connected. Since a similar argument works for $S^{-1}(E)$, the proof is complete.

Acknowledgment. We are pleased to thank Erwin Kronheimer for many insightful comments.

REFERENCES

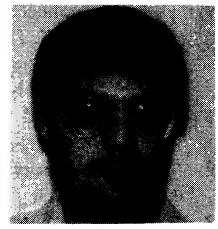
1. P. Alexandroff, *Diskrete Räume*, *Mat. Sbornik*, 2 (1937) 501–518.
2. Peter T. Johnstone, Stone Spaces, Cambridge Univ. Press, Cambridge, 1982.
3. J. L. Kelley, General Topology, New York, 1955.
4. E. D. Khalimsky (E. Halimskii), On topologies of generalized segments, *Soviet Math. Doklady*, 10 (1969) 1508–1511.
5. E. D. Khalimsky, Applications of connected ordered topological spaces in topology, Conference of math. departments of Povolsia, 1970.

6. _____, *Uporiadochenie Topologicheskie Prosftranstva* (*Ordered Topological Spaces*), Naukova Dumka Press, Kiev, 1977.
7. E. D. Khalimsky, R. Kopperman, P. R. Meyer, Computer graphics and connected topologies on finite ordered sets, *Topology and its Applications*, 36 (1990) 1–17.
8. E. D. Khalimsky, R. Kopperman, P. R. Meyer, Boundaries in digital planes, *Journal of Applied Math. and Stocastic Analysis* 3 (1990) 27–55.
9. H. Kok, Connected Orderable Spaces, Math. Centrum, Amsterdam, 1973.
10. T. Y. Kong, E. D. Khalimsky, Polyhedral analogs of locally finite topological spaces, in: R. M. Shortt, Ed., General Topology and Applications, Proceedings of the 1988 Northeast Conference, Marcel Dekker, NY, 1990, 153–164.
11. T. Y. Kong and A. W. Roscoe, Continuous analogs of axiomatized digital surfaces, *Computer Vision, Graphics, and Image Processing*, 29 (1985) 60–86.
12. _____, A theory of binary digital pictures, *Computer Vision, Graphics, and Image Processing*, 32 (1985) 221–243.
13. T. Y. Kong, A. Rosenfeld, Digital topology: Introduction and survey, *Computer Vision, Graphics, and Image Processing*, 48 (1989) 357–393.
14. R. Kopperman, P. R. Meyer, R. G. Wilson, A Jordan surface theorem for three-dimensional digital spaces, *Discrete and Computational Geometry*, 6 (1991) 155–161.
15. V. A. Kovalevsky, On the topology of discrete spaces, Studientexte, Digitale Bildverarbeitung, Heft 93/86, Technische Universitat Dresden 1986.
16. _____, Finite topology as applied to image analysis, *Computer Vision, Graphics, and Image Processing*, 46 (1989) 141–161.
17. E. H. Kronheimer, The topology of digital images, *Topology and its Applications*, in press.
18. D. Marcus, A special topology for the integers, *Amer. Math. Monthly*, 77 (1970) 1119.
19. P. R. Meyer, Topologies of separate continuity in a cross space, *Amer. Math. Monthly*, 75 (1968) 1128.
20. G. M. Reed, On the characterization of simple closed surfaces in three-dimensional digital images, *Computer vision, Graphics, and Image Processing*, 25 (1984) 226–235.
21. G. M. Reed and A. Rosenfeld, Recognition of surfaces in three-dimensional digital images, *Information and Control*, 53 (1982) 108–120.
22. A. Rosenfeld, Digital topology, *Amer. Math. Monthly*, 86 (1979) 621–630.
23. _____, Picture Languages, Academic Press, NY, 1979.

Linear Algebra by Analogy

SCOTT H. HOCHWALD, *University of North Florida, Jacksonville, FL 32216*

SCOTT H. HOCHWALD: I discovered the beauty of mathematics while I was a member of a Brooklyn Tech High School class that spent three years under the guidance of Lawrence J. Zimmerman. I learned how to do mathematics at Caltech, where Richard Dean and Fred Dashiell played key roles in that phase of my education. I received my Ph.D. in mathematics from U.C. Berkeley in 1984 under the direction of Donald Sarason. Since that time, I have been on the faculty of the University of North Florida. My main research interests are operator theory and linear algebra.



1. Introduction. Whenever I see a result about real or complex numbers, I ask myself whether it is also true for linear transformations. I developed this habit as a result of reading books like [**9**] and articles like [**6**].

One reason I wrote this article is so that others may pick up this habit. Another reason I wrote it is to emphasize aspects of linear algebra that suffer from underexposure.

My use of the phrase "linear transformations" as opposed to the word "matrices" is deliberate. A great deal can be said about linear algebra without appealing to matrices. In fact, except for 2×2 matrix examples and counterexamples, I could have written this article without mentioning matrices. I chose not to do this so that insights were not obscured by techniques. Those who want to fill in any coordinate-free gaps or who want to see a coordinate-free treatment of linear algebra should read [**9**].

In order to make this article accessible to as large an audience as possible, the underlying vector space is restricted to being finite-dimensional. Much of what is true in the finite-dimensional case is also true in the infinite-dimensional (complete) case, and what isn't true is what makes the study of infinite-dimensional spaces interesting. Therefore, I'll occasionally note some of the complications that occur in infinite-dimensions.

2. Introducing the main players. Let $L(\mathbf{V})$ denote the set of all linear transformations on a complex finite-dimensional inner product space $\mathbf{V}$ of dimension n, where $n > 1$. Let $\mathbb{C}$ denote the set of complex numbers and $(\cdot, \cdot)$ denote the inner product on $\mathbf{V}$.

In what follows, all 2×2 matrices are written with respect to the orthonormal basis $\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$ for the inner product space consisting of the set of all ordered pairs of complex numbers with the dot product.

If $S \in L(\mathbf{V})$, then S^*, the *adjoint* of S, is the unique linear transformation that satisfies $(Sx, y) = (x, S^*y)$ for all x, y in $\mathbf{V}$. If one fixes an orthonormal basis for $\mathbf{V}$ then the matrix of S^* with respect to this basis is the conjugate transpose of the matrix of S with respect to it. It follows directly from either of the last two statements that for S, T in $L(\mathbf{V})$: $(S + T)^* = S^* + T^*$, $(ST)^* = T^*S^*$, $(S^*)^* = S$, and $(kS)^* = \bar{k}S^*$ if $k \in \mathbb{C}$.

In an algebraic sense, $L(\mathbf{V})$ with the adjoint operation looks very much like $\mathbb{C}$ with the conjugation operation. (This is the raison d'être of this article.) Their

algebraic differences occur because multiplication in $L(\mathbf{V})$ is not commutative and there are nonzero linear transformations that do not have multiplicative inverses. One of these conditions has a much greater impact than the other. Care to make a conjecture?

Given any complex number z, there exist unique real numbers a and b such that $z = a + bi$. In order to transfer this basic observation regarding complex numbers into one concerning linear transformations, the "real" linear transformations must be identified. Since the complex number z is real if and only if $z = \bar{z}$, we say a linear transformation S is *Hermitian* ("real") if $S = S^*$. Since

$$\mathrm{Re}(z) = \frac{z + \bar{z}}{2} \text{ and } \mathrm{Im}(z) = \frac{z - \bar{z}}{2i} \text{ for a complex number } z,$$

we define the real and imaginary parts of a linear transformation S by

$$Re(S) = \frac{S + S^*}{2} \text{ and } Im(S) = \frac{S - S^*}{2i}.$$

It then follows that both $\mathrm{Re}(S)$ and $\mathrm{Im}(S)$ are Hermitian and $S = \mathrm{Re}(S) + i\,\mathrm{Im}(S)$.

If S is Hermitian, then (Sx, x) is real for all x in V, because $(Sx, x) = (x, Sx) = \overline{(Sx, x)}$. The converse of this observation is also true. That is, if (Sx, x) is real for all x in $\mathbf{V}$, then S is Hermitian ([**9**, p. 139]).

It is easy to see that the set of Hermitian linear transformations, like the set of real numbers, is closed under addition. However, the situation is different for multiplication. For example, if

$$S = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \text{ and } T = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix},$$

then S and T are both Hermitian, but $ST = \begin{pmatrix} 0 & -4 \\ 1 & 9 \end{pmatrix}$ is not. It is because multiplication is not commutative that we get this counterexample. If S and T are Hermitian linear transformations and S commutes with T, then ST is Hermitian, simply because $(ST)^* = (TS)^* = S^*T^* = ST$.

Now that the "real" linear transformations have been identified, it follows naturally that we look for the "pure imaginary" linear transformations, the "positive" linear transformations, and the "negative" linear transformations.

Since a complex number z is pure imaginary if and only if $z = -\bar{z}$, we say that a linear transformation S is *skew-Hermitian* ("pure imaginary") if $S = -S^*$.

A moment's thought about connections between real and pure imaginary numbers and a few simple manipulations lead us to conclude that a linear transformation S is skew-Hermitian if and only if iS is Hermitian. Thus, the set of skew-Hermitian linear transformations is closed under addition and, as expected, it is not closed under multiplication.

The definition of a "positive" linear transformation is not as clear cut. It is best approached by first defining a "nonnegative" linear transformation, although this also has its complications. On the one hand, since a real number p is nonnegative if and only if $p = \bar{z}z$ for some complex number z, we should say that a Hermitian linear transformation S is *nonnegative* if $S = T^*T$ for some linear transformation T. On the other hand, since a real number p is nonnegative if and only if $p = b^2$ for some real number b, we should say that a Hermitian linear transformation S is *nonnegative* if $S = R^2$ for some Hermitian linear transformation R.

It turns out that it is not necessary to decide which definition is better because they are equivalent. (One way to show this is to use the Diagonalization Theorem in section 4.)

If $S \in L(\mathbf{V})$ and $S = T^*T$ for some T in $L(\mathbf{V})$, then S is Hermitian. So, it makes sense to simply say, "S is a nonnegative linear transformation," and to denote this by $S \geq 0$.

A linear transformation S is said to be *nonpositive* if $-S \geq 0$.

If S and T are Hermitian linear transformations and $S - T \geq 0$, then we write $S \geq T$. This natural partial order turns out not to be a total order. The following proposition and a "natural counterexample" illustrate why this is the case. In addition, the proposition establishes a direct connection between "$\geq$" for nonnegative real numbers and "$\geq$" for nonnegative linear transformations.

PROPOSITION 2.1. *The linear transformation S is nonnegative if and only if $(Sx, x) \geq 0$ for all x in* **V**.

See [**9**, p. 166] for a proof of the hard direction.

If

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

then $(Sx, x) = 1$ and $(Sy, y) = -1$. It then follows from Proposition 2.1 that $S \not\geq 0$ and $0 \not\geq S$. Thus, unlike the set of real numbers, the set of Hermitian linear transformations cannot be totally ordered by the natural partial order.

Another immediate consequence of Proposition 2.1 is that the set of nonnegative linear transformations is closed under addition.

Again, the situation is different for multiplication. For example, if

$$S = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

and

$$T = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix},$$

then $ST = \begin{pmatrix} 0 & -4 \\ 1 & 9 \end{pmatrix}$. This shows that the set of nonnegative linear transformations is not closed under multiplication. Once again, this is a consequence of the fact that multiplication is not commutative. That is, if two nonnegative linear transformations commute, then their product is nonnegative ([**9**, p. 167]).

In view of the discussion concerning nonnegative linear transformations, it seems reasonable to say that a linear transformation, S, is *positive*, in symbols $S > 0$, if $(Sx, x) > 0$ for all x in V. *If S is positive, then S is invertible and nonnegative. This implies that $S = T^*T$ for some invertible linear transformation T and $S = R^2$ for some invertible Hermitian linear transformation R.* These two sentences are in italics because even when "properly modified" they are not valid when **V** is infinite dimensional. "Properly modified" means that the phrase "linear transformation" is replaced by the phrase "bounded linear transformation" (see [**9**, p. 176] for a definition).

The reason for starting the above discussion by identifying the nonnegative linear transformations rather than the positive linear transformations can now be revealed. A nonnegative real number is positive if it is nonzero; this is equivalent

to its being invertible (with respect to multiplication). However, these two terms, nonzero and invertible, are not equivalent when used to describe linear transformations. Thus, the translation of the statement, "a real number p is positive if there exists a nonzero complex number z such that $p = \bar{z}z$" to a statement about linear transformations is ambiguous. It now appears that "nonzero complex number" corresponds to "invertible linear transformation."

The final subset of the complex plane that has a role in the analogies is the unit circle. A convenient way to describe this set so that the corresponding set of linear transformations can be identified is $\{z\colon z \in \mathbb{C} \text{ and } \bar{z}z = 1\}$. Thus, we say that an invertible linear transformation U is *unitary* if $U^*U = I$. (This is not redundant when V is infinite-dimensional.) If U and V are unitary linear transformations, then $(UV)^*UV = V^*U^*UV = I$. Hence, the set of unitary linear transformations, like the unit circle, is closed under multiplication.

3. Natural questions. This section contains answers to many questions that arise naturally when $L(\mathbf{V})$ is viewed in the light of the previous section. Most of the results in this section just appear as statements of theorems. To fully appreciate these results their proofs must be sampled.

The following theorem demonstrates that nonnegative linear transformations behave like nonnegative real numbers with respect to taking roots.

ROOT THEOREM. *If S is a nonnegative linear transformation and n is a positive integer, then there exists a unique nonnegative linear transformation T such that $T^n = S$ (so $S^{1/n} = T$).*

See [3] for a proof. (Another way to prove this is to use the Diagonalization Theorem in section 4.)

If z is a complex number, then there is a unique nonnegative real number r ($r = |z| = (\bar{z}z)^{1/2}$) and a real number θ in $[0, 2\pi)$ such that $z = re^{i\theta}$. If z is nonzero, then r is positive and θ is also uniquely determined.

The next analogous result places a different emphasis upon this decomposition.

POLAR DECOMPOSITION THEOREM ([15], [9, p. 169]). *If $S \in L(\mathbf{V})$, then there is a unique nonnegative linear transformation P, where $P = (S^*S)^{1/2}$, and a unitary linear transformation U such that $S = UP$. If S is invertible, then P is positive and U is also uniquely determined.*

There is more to say about this decomposition but the statements and details must wait until section 4.

If $S \in L(V)$, let $|S|$ denote $(S^*S)^{1/2}$. The next five results need no added explanation.

THEOREM 3.1 ([5]). *If $S \in L(\mathbf{V})$ and $(\mathrm{Re}(S))^2 \geq |S|^2$, then S is Hermitian.*

THEOREM 3.2 ([6]). *If $S \in L(\mathbf{V})$ and $\mathrm{Re}(S) \geq |S|$, then S is nonnegative.*

THEOREM 3.3 ([6]). *If $S \in L(\mathbf{V})$ and $|\mathrm{Re}(S)| \geq |S|$, then S is Hermitian.*

THEOREM 3.4 ([14]). *If $S, T \in L(\mathbf{V})$ and $S \geq T \geq 0$, then $S^{1/2} \geq T^{1/2} \geq 0$.*

THEOREM 3.5 ([1]). *If $S, T \in L(\mathbf{V})$, T is invertible, and $S \geq T \geq 0$, then S is invertible and $T^{-1} \geq S^{-1} \geq 0$.*

The "properly modified" version of Theorem 3.3 is unresolved in the infinite-dimensional case.

Nice proofs of Theorems 3.4 and 3.5 and much more in the spirit of this article can be found in [4]. Those interested in [4] may also be interested in [7] and [8].

We now break the trend established by the last few theorems and show that some inequalities about real numbers cannot be translated to inequalities about Hermitian linear transformations.

There exist linear transformations S and T such that $S \geq T \geq 0$, but $S^2 \not\geq T^2$. Let $S = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ and $T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. It is easy to see that $T \geq 0$, and we already know that $S \geq 0$. Since

$$S - T = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix},$$

it follows that $S \geq T$. Since

$$S^2 - T^2 = \begin{pmatrix} 2 & -3 \\ -3 & 4 \end{pmatrix}$$

and $((S^2 - T^2)x, x) = -.5$ when $x = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$, it follows that $S^2 \not\geq T^2$.

The story is different if S and T commute. If $S \geq T \geq 0$ and S commutes with T, then $S^2 \geq T^2$. Here are the details. Since both $S + T$ and $S - T$ are nonnegative and they commute with each other, it follows that $S^2 - T^2 = (S + T)(S - T) \geq 0$.

There exist Hermitian linear transformations S and T such that $S \geq 0$ and $S \geq T \geq -S$ but $S \not\geq |T|$. Let

$$S = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & \sqrt{2} \\ 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \sqrt{2} & \sqrt{2} \end{pmatrix}$$

and $T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. It then follows that

$$S - T = \begin{pmatrix} \sqrt{2} & 0 \\ \sqrt{2} & 1 \end{pmatrix}\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 1 \end{pmatrix}, \qquad S + T = \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix},$$

but

$$S - |T| = \begin{pmatrix} 2 & 2 \\ 2 & 1 \end{pmatrix} \text{ and } ((S - |T|)x, x) = -1 \text{ when } x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Once again, if S and T commute, then $S \geq 0$ and $S \geq T \geq -S$ imply that $S \geq |T|$. This follows from the facts that $S^2 \geq T^2$ and $S = (S^2)^{1/2} \geq (T^2)^{1/2} = |T|$.

4. Normal linear transformations. This brief section explores another important effect of commutativity and completes a discussion that began in section 3. Even though the following sets are characterized by the commutativity of different linear transformations, they are equal.

$A = \{S \in L(\mathbf{V})$: S commutes with $S^*\}$

$B = \{S \in L(\mathbf{V})$: Re$(S)$ commutes with Im$(S)\}$

$C = \{S \in L(\mathbf{V})$: $S = UP$, a Polar Decomposition, and U commutes with $P\}$

A few simple calculations show that $A = B$. A couple of different calculations show that $C \subseteq A$. The following shows that $A \subseteq C$. If $SS^* = S^*S$ and $S = UP$ is a

Polar Decomposition of S, then $UPPU^* = PU^*UP$. This shows that $UP^2 = P^2U$. Since U commutes with P^2, it follows that U commutes with P, the unique positive square root of P^2. See [**9**, p. 170] for more details.

It is easy to see that

$$D = \left\{S \in L(V): (\mathrm{Re}(S))^2 + (\mathrm{Im}(S))^2 = |S|^2\right\}$$

also belongs on the list of sets.

We say that a linear transformation S is *normal* if $SS^* = S^*S$. Nonnegative, Hermitian, skew-Hermitian, and unitary linear transformations are normal.

The following fact is even more remarkable than the equality of A, B, C, and D.

DIAGONALIZATION THEOREM ([**13**, p. 304]). *The linear transformation S is normal if and only if there exists an orthonormal basis for* $\mathbf{V}$ *consisting of eigenvectors of S.*

We are now able to complete the discussion concerning Polar Decompositions. Basically, we show that unitary linear transformations have a form analogous to the "$e^{i\theta}$" form for complex numbers on the unit circle.

If T is a normal linear transformation, then there exists an orthonormal basis $\{v_j\}$ for V and complex numbers λ_j such that $Tv_j = \lambda_j v_j$ for $j = 1, 2, \ldots, n$. The linear transformation e^T is defined by $e^T v_j = e^{\lambda_j} v_j$.

It is actually true that e^T makes good sense even when T is not normal. Formally, $e^T = I + \sum_{n=1}^{\infty} T^n/n!$. It can be shown that this infinite series converges to a linear transformation and, moreover, the two definitions coincide when T is normal.

If U is a unitary linear transformation and x is an eigenvector of U, then $(Ux, Ux) = (U^*Ux, x) = (x, x) = 1$ shows that the set of eigenvalues of U is a subset of the unit circle.

It is now easy to prove the following result.

THEOREM 4.1. *If U is a unitary linear transformation, then there exists a Hermitian linear transformation T such that $U = e^{iT}$.*

Theorem 4.1 remains valid in the infinite-dimensional case (see [**16**, p. 317]) as long as it is properly modified. However, the Polar Decomposition Theorem must be more than properly modified to remain valid in the infinite-dimensional case (see [**10**, p. 74]).

5. The spectrum and the trace. I'm sure that it is now apparent that the source of a lot of the interesting algebraic phenomena in $L(\mathbf{V})$ is commutativity—or the lack thereof. This point of view has some surprising consequences when applied to the *spectrum* and the *trace*.

If $S \in L(\mathbf{V})$, let *spec*(S), the *spectrum* of S, denote the set of eigenvalues of S and let *tr*(S), the *trace* of S, denote the sum of the eigenvalues (counting multiplicities) of S. Equivalently, tr(S) = the sum of the diagonal entries of any matrix representation of S.

We begin the investigation by concentrating on spec$(\cdot)$. Let spec$(S)^*$ denote $\{\bar{\lambda}: \lambda \in \mathrm{spec}(S)\}$ and if $\underline{k} \in \mathbb{C}$, let $\underline{k}\,\mathrm{spec}(S)$ denote $\{\underline{k}\lambda: \lambda \in \mathrm{spec}(S)\}$.

If $\lambda \in \mathrm{spec}(S)$, then there exists a unit vector x such that $(Sx, x) = (\lambda x, x) = \lambda(x, x) = \lambda$. This implies that $\mathrm{spec}(S) \subseteq [0, \infty)$ if S is nonnegative, spec$(S) \subseteq$

$(-\infty, \infty)$ if S is Hermitian, and $\text{spec}(S) \subseteq \{\text{pure imaginary numbers}\}$ if S is skew-Hermitian.

It is also easy to see that if $\underline{k} \in \mathbb{C}$, then $\text{spec}(\underline{k}S) = \underline{k}\,\text{spec}(S)$.

Another satisfying result concerns the relationship between $\text{spec}(S)$ and $\text{spec}(S^*)$.

PROPOSITION 5.1. *If $S \in L(\mathbf{V})$, then $spec(S^*) = spec(S)^*$.*

See [**12**, p. 57] for a proof.

Another more remarkable result along these lines concerns products.

PROPOSITION 5.2. *If S and T are in $L(\mathbf{V})$, then $spec(ST) = spec(TS)$.*

The details for this can be found in [**17**]. This statement needs slightly more than to be properly modified when $\mathbf{V}$ is infinite dimensional. See [**10**, p. 43] for details.

Proposition 5.2 shows that, in some sense, $\text{spec}(\cdot)$ allows any two linear transformations to commute. Here are some of the consequences of this.

THEOREM 5.1. *If S and T are nonnegative linear transformations, then $spec(ST) \subseteq [0, \infty)$.*

Proof. Since $S \geq 0$, there exists a nonnegative linear transformation R such that $R^2 = S$. It then follows that $\text{spec}(ST) = \text{spec}(R^2T) = \text{spec}(RTR)$. If $x \in \mathbf{V}$, then $(RTRx, x) = (TRx, Rx) = (Ty, y) \geq 0$. Therefore, $RTR \geq 0$ and thus, $\text{spec}(RTR) \subseteq [0, \infty)$. ■

For a different proof of Theorem 5.1 and interesting related results see [**18**]. The proof in [**18**] must, however, be modified so that it works for noninvertible linear transformations.

Even though it is true that the product of commuting Hermitian linear transformations is Hermitian, it is not possible to replace "nonnegative" in Theorem 5.1 by "Hermitian" and "$[0, \infty)$" by "$(-\infty, \infty)$". Consider for example, $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. It is easy to see that S and T are Hermitian, $ST = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $\text{spec}(ST) = \{i, -i\}$.

Nevertheless, there is something salvageable from this situation. If S and T are Hermitian linear transformations, then $\text{spec}(ST) = \text{spec}(TS) = \text{spec}((ST)^*)$. This implies that $\lambda \in \text{spec}(ST)$ if and only if $\bar{\lambda} \in \text{spec}(ST)$. This is equivalent to saying that $\text{spec}(ST)$ is symmetric about the real axis.

Notice that $\text{spec}(ST) \subseteq (-\infty, \infty)$ if one of S or T is Hermitian and the other is nonnegative.

Can Theorem 5.1 be extended from two factors to three? No. Let

$$R = \begin{pmatrix} 2 & -i \\ i & 1 \end{pmatrix} = \begin{pmatrix} 1 & -i \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ i & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad \text{and } T = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}.$$

It then follows that R, S, and T are nonnegative, $RST = \begin{pmatrix} -i & -8-9i \\ 1 & 9-4i \end{pmatrix}$, and $\text{spec}(RST)$ contains two complex numbers whose sum is $9 - 5i$. We expect this because, unlike the situation for two linear transformations, $\text{spec}(\cdot)$ "doesn't allow" three linear transformations to mutually commute. In fact, the above example shows that $\text{spec}(TSR)$ contains two complex numbers whose sum is $9 + 5i$ ($\text{spec}(TSR) = \text{spec}(RST)^*$ since $TSR = (RST)^*$).

A positive result along these lines is: If R, S, and T are nonnegative linear transformations and RST is Hermitian, then $\operatorname{spec}(RST) \subseteq [0, \infty)$. In fact, these conditions imply that RST is nonnegative. See [**18**] for details.

The time has come to let $\operatorname{tr}(\cdot)$ enter the picture.

Clearly, $\operatorname{tr}(S) \geq 0$ if S is nonnegative, $\operatorname{tr}(S) \in (-\infty, \infty)$ if S is Hermitian, and $\operatorname{tr}(S) \in \{\text{pure imaginary numbers}\}$ if S is skew-Hermitian.

In light of what has been mentioned concerning $\operatorname{spec}(\cdot)$ and/or by definition it is now easy to see that if S and T are linear transformations and k is a complex number, then $\operatorname{tr}(kS) = k \operatorname{tr}(S)$, and $\operatorname{tr}(S^*) = \overline{\operatorname{tr}(S)}$, $\operatorname{tr}(ST) = \operatorname{tr}(TS)$, and if in addition, S and T are nonnegative, then $\operatorname{tr}(ST) \geq 0$.

A perusal of the beginning of this section shows that we have a perfect match in results up to this point. The first discrepancy occurs when we consider $\operatorname{tr}(ST)$ for Hermitian S and T. In this case, $\operatorname{tr}(ST) \in (-\infty, \infty)$ but all that can be said for $\operatorname{spec}(ST)$ is that it is symmetric about the real axis.

Formally, we can account for this "better" behavior of $\operatorname{tr}(\cdot)$ with a proof. If $S = [s_{ij}]$ and $T = [t_{ij}]$ are matrix representations for S and T with respect to the same basis for V, then $s_{ij} = \bar{s}_{ji}$, $t_{ij} = \bar{t}_{ji}$ and the sum of the diagonal elements of ST is $\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}t_{ji}$. The following calculations show that this sum is indeed a real number.

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}t_{ji} &= \sum_{i=2}^{n}\sum_{j=1}^{i-1} s_{ij}t_{ji} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} s_{ij}t_{ji} + \sum_{i=1}^{n} s_{ii}t_{ii} \\
&= \sum_{i=2}^{n}\sum_{j=1}^{i-1} s_{ij}\bar{t}_{ij} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \bar{s}_{ji}t_{ji} + \sum_{i=1}^{n} s_{ii}t_{ii} \\
&= \sum_{i=2}^{n}\sum_{j=1}^{i-1} s_{ij}\bar{t}_{ij} + \sum_{j=2}^{n}\sum_{i=1}^{j-1} \bar{s}_{ji}t_{ji} + \sum_{i=1}^{n} s_{ii}t_{ii}.
\end{aligned}$$

The last expression is the sum of a complex number, its conjugate and a real number.

Actually, more is true. There is, in fact, a basis for V such that the matrix representation of ST with respect to this basis consists only of real numbers. See [**12**, Theorem 4.1.7] for the details.

Informally, additivity might explain this "better" behavior. That is, $\operatorname{tr}(S + T) = \operatorname{tr}(S) + \operatorname{tr}(T)$ and it seems that it is not possible for $\operatorname{spec}(\cdot)$ to be additive. However, if A and B are sets of complex numbers, then $A + B = \{a + b\colon a \in A \text{ and } b \in B\}$ defines addition for sets. Clearly, $\operatorname{spec}(S + T) = \operatorname{spec}(S) + \operatorname{spec}(T)$ is too much to hope for. Is $\operatorname{spec}(S) + \operatorname{spec}(T) \subseteq \operatorname{spec}(S + T)$ or $\operatorname{spec}(S + T) \subseteq \operatorname{spec}(S) + \operatorname{spec}(T)$? No. If $S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, then $\operatorname{spec}(S) = \operatorname{spec}(T) = \{0\}$ and $\operatorname{spec}(S + T) = \{-1, 1\}$. Once again, the situation changes when commutativity enters the picture. That is, if S and T commute, then $\operatorname{spec}(S + T) \subseteq \operatorname{spec}(S) + \operatorname{spec}(T)$. See [**12**, p. 92] for the details. Thus, $\operatorname{tr}(\cdot)$ is additive and $\operatorname{spec}(\cdot)$ isn't.

Final remarks. Any discussion of linear algebra that mentions the trace usually puts a greater emphasis on the determinant. Therefore, I feel that I must justify the fact that section 5 has a discussion of the trace without mentioning the determinant.

Since $\det(ST) = \det(S)\det(T)$, all the results in section 5 become trivial when stated properly for the determinant. However, since $\det(\cdot)$ is not additive, we should investigate $\det(S + T)$. Once again, results become trivial because $S + T$ is nonnegative when S and T are nonnegative and $S + T$ is Hermitian when S and T are Hermitian.

It is possible to restate Theorem 5.1 in a way that raises an interesting question. Namely, if S is a nonnegative linear transformation, then $\operatorname{spec}(ST) \subseteq [0, \infty)$ for all nonnegative linear transformations T.

Is the converse of this theorem true? That is, if $\operatorname{spec}(ST) \subseteq [0, \infty)$ for all nonnegative linear transformations T, does it follow that S is nonnegative? The answer is yes and it is a consequence of Theorem 4 in [**2**]. Also, [**11**] has an approach to this question and related questions that is different from the approach in [**2**].

Acknowledgement. I thank my UNF colleagues Marsha Finkel-Babadi and Leonard Lipkin, and Lawrence J. Zimmerman for their suggestions on improving the exposition of one of the earlier versions of this article.

I especially thank Paul R. Halmos for his comments and for pointing out the rough spots in the original version and in a subsequent revision.

REFERENCES

1. Y. H. Au-Yeung, Some inequalities for the rational powers of a nonnegative definite matrix, *Lin. Alg. and Appl.*, 7 (1973) 347–350.
2. C. S. Ballantine and R. Loewy, Self-duality results for unitary and Hermitian matrices, *Lin. and Multilin. Alg.*, 13 (1983) 221–230.
3. D. R. Brown and M. J. O'Malley, On nth roots of positive operators, this MONTHLY, 87 (1980) 380–382.
4. N. N. Chan and M. K. Kwong, Hermitian matrix inequalities and a conjecture, this MONTHLY, 92 (1985) 533–541.
5. C. K. Fong and V. I. Istratescu, Some characterizations of Hermitian operators and related classes of operators. I, *Proc. Amer. Math. Soc.*, 76 (1979) 107–112.
6. C. K. Fong and S. K. Tsui, A note on positive operators, *J. Operator Theory*, 5 (1981) 73–76.
7. T. Furuta, A counterexample to a conjectured Hermitian matrix inequality, this MONTHLY, 94 (1987) 271–272.
8. _____, $A \geq B \geq 0$ ensures $(BA^2B)^{1/2} \geq B^2$ – solution to a conjecture on operator inequalities, this MONTHLY, 95 (1988) 933–934.
9. P. R. Halmos, Finite-Dimensional Vector Spaces, Springer-Verlag, N.Y., 1974.
10. _____, A Hilbert Space Problem Book, Springer-Verlag, N.Y., 1982.
11. S. H. Hochwald, Natural spectral conditions on operators (in preparation).
12. R. A. Horn and C. A. Johnson, Matrix Analysis, Cambridge University Press, N.Y., 1985.
13. S. Lipschutz, Schaums Outline of Theory and Problems of Linear Algebra, McGraw-Hill, N.Y., 1968.
14. C. Loewner, Über monotone Matrixfunktionen, *Math Z.*, 38 (1934) 177–216.
15. L. C. Martins and P. Podio-Guidugli, An elementary proof of the polar decomposition theorem, this MONTHLY, 87 (1980) 288–290.
16. W. Rudin, Functional Analysis, McGraw-Hill, N.Y., 1973.
17. _____, Unique right inverses are two-sided, this MONTHLY, 92 (1985) 489–490.
18. E. P. Wigner, On weakly positive matrices, *Canad. J. Math.*, 15 (1963) 313–317.

NOTES

EDITED BY DENNIS DETURCK, DAVID HALLENBECK, AND RODICA SIMION

An Obvious Proof of Burnside's Lemma

KENNETH P. BOGART*
Department of Mathematics and Computer Science, Dartmouth College, Hanover, NH 03755

The Burnside-Cauchy-Frobenius (B-C-F) formula [2, 4], which is commonly called Burnside's lemma, allows us to compute the number of orbits (equivalence classes) induced by a group G of permutations on a set X (where x and y are equivalent if $gx = y$ for some g in G). The usual proof is an elementary computation involving the relationship between an orbit and the cosets of the subgroup fixing one of its elements [1, 2, 3]. The purpose of this note is to give a proof which makes the result an obvious corollary of the product principle.

The product principle for sets says that the union of n disjoint sets each of size m has $m \cdot n$ elements. For multisets the product principle says that the multiset union of n disjoint multisets each of size m is a multiset of size mn.

The intuitive idea that a multiset is a set with multiplicities and its size is the sum of the multiplicities should suffice for reading this paper. It is, however, worth noting that a function f from a set S to a set T defines a multiset of size $|S|$ chosen from T in which the multiplicity of $t \in T$ is the number of $s \in S$ with $f(s) = t$, i.e., $|f^{-1}(\{t\})|$. In particular, given a group G of permutations of a set X and an element x of X, the function $g \to gx$ from G to X determines a multiset denoted by

$$Gx = [gx | g \in G]$$

of size $|G|$ chosen from X. We call this multiset a multiorbit of x. Observe that, as with ordinary orbits, two multiorbits Gx and Gy are either identical or disjoint. As a consequence we have the following Lemma.

LEMMA. *The multiplicity of an element x in a multiorbit is the number of elements of G that fix x.*

Proof. If x is in Gy, then the multisets Gy and Gx are equal. Clearly the multiplicity of x in Gx is the number of elements that fix x. ■

By the product principle, if there are n distinct multiorbits, then their union has size $n|G|$. The size of this union is the sum of the multiplicities of the elements of X in this union. Since each member lies in one and only one multiorbit, we obtain

$$\begin{aligned} n|G| &= \sum_{x:\, x \in X} |\{g:\, gx = x\}| \\ &= |\{(x, g):\, gx = x\}| \\ &= \sum_{g:\, g \in G} |\{x:\, gx = x\}|. \end{aligned}$$

*Work supported by ONR Contract N00014-88-K-0065

Poisson Integrals of Riemann Integrable Functions

J. KRÁL
MU ČSAV, Žitná 25, 115 67 Praha 1, Czechoslovakia

W. F. PFEFFER
Department of Mathematics, University of California, Davis, CA 95616

Dedicated to Professor Jan Mařík for his 70th birthday

Throughout, $\mathbf{R}$ and $\mathbf{C}$ denote, respectively, the sets of all real and complex numbers. By a function, we always mean a complex-valued function. We let

$$D = \{z \in \mathbf{C}: |z| < 1\} \quad \text{and} \quad T = \{z \in \mathbf{C}: |z| = 1\}.$$

An infinitely differentiable function u defined in an open set $\Omega \subset \mathscr{E}$ is called *harmonic* in Ω if it satisfies the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

at each point $z = (x, y)$ of Ω. The *Poisson kernel* is the positive function

$$P_r(t) = \frac{1}{2\pi} \cdot \frac{1 - r^2}{1 - 2r\cos t + r^2}$$

defined for $0 \leq r < 1$ and $t \in \mathbf{R}$. If f is a Lebesgue integrable function defined in $[-\pi, \pi]$, then

$$u(re^{i\theta}) = \int_{-\pi}^{\pi} P_r(\theta - t) f(t)\, dt,$$

where $0 \leq r < 1$ and $\theta \in \mathbf{R}$, defines a harmonic function u in D, denoted by $P * f$ (see [**6**; Theorem 11.7, p. 234]). Moreover, if f is continuous and $f(\pi) = f(-\pi)$, then $P * f$ extends continuously to $D \cup T$ and $P * f \restriction T = f$. In this case, $P * f$ is the unique solution of the following *Dirichlet problem*: find a continuous function defined in $D \cup T$ which is harmonic in D and whose restriction to T equals f (cf. [**2**]).

Thus given a class $\mathscr{E}$ of Lebesgue integrable functions defined in $[-\pi, \pi]$, it is a natural task to characterize the functions u harmonic in D such that $u = P * f$ for an $f \in \mathscr{E}$. When $\mathscr{E} = \mathscr{L}^p$, such characterizations are well known (see [**6**, Theorems 11.16 and 11.30, pp. 239 and 247] if $1 < p \leq \infty$, and [**3**, Chapter I, Section 3.2, p. 43] if $p = 1$). In this note we give a simple characterization for the class $\mathscr{R}$ of all *Riemann integrable* functions defined in $[-\pi, \pi]$. The proof, which is elementary and reasonably self-contained, is based on the classical result that a bounded function is Riemann integrable if and only if it is continuous almost everywhere (see [**5**, Theorem 11.33(b), p. 323]).

A direct calculation shows that

$$\int_{\theta_0 - \pi}^{\theta_0 + \pi} P_r(\theta - t)\, dt = 1 \tag{1}$$

for each $\theta, \theta_0 \in \mathbf{R}$ and $0 \leq r < 1$. As every Riemann integrable function f defined in $[-\pi, \pi]$ is bounded by definition, if follows from (1) that $u = P * f$ is bounded

in D. Thus it suffices to consider only the class h^∞ of all bounded harmonic functions in D (cf. [**1**; Section 1.1, p. 2]).

If $E \subset (-\pi, \pi)$, we denote by $\mathscr{C}_E$ the family of all bounded measurable functions defined in $[-\pi, \pi]$ which are continuous at each point $t \in E$.

PROPOSITION. *Let $u \in h^\infty$, and let $u_r(\theta) = u(re^{i\theta})$ for $0 \leq r < 1$ and $\theta \in \mathbf{R}$. There is an $f \in \mathscr{C}_E$ with $P * f = u$ if and only if the family $\{u_r\colon 0 \leq r < 1\}$ is equicontinuous at each $\theta \in E$.*

Recall that the family $\{u_r\colon 0 \leq r < 1\}$ is *equicontinuous* at $\theta_0 \in \mathbf{R}$ whenever given $\varepsilon > 0$, there is a $\delta > 0$ such that $|u_r(\theta) - u_r(\theta_0)| < \varepsilon$ for each $r \in [0, 1)$ and $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$.

Proof. Suppose first that $u = P * f$ for an $f \in \mathscr{C}_E$. If $\theta \in E$ then $\lim_{r \to 1-} u_r(\theta) = f(\theta)$ by [**5**; Theorem 11.23, p. 244], and the equicontinuity of the family $\{u_r\colon 0 \leqslant r < 1\}$ at θ follows. Nonetheless, for completeness we shall prove this directly. With no loss of generality, we may assume that $f(\pi) = f(-\pi)$ and view f as periodically extended to $\mathbf{R}$ with period 2π. Choose an $a > 0$ with $|f(t)| \leqslant a$ for each $t \in \mathbf{R}$. Given a $\theta_0 \in E$ and $\varepsilon > 0$, find a positive $\delta' \leqslant \pi/2$ so that $|f(t) - f(\theta_0)| < \varepsilon/4$ for each $t \in \mathbf{R}$ with $|t - \theta_0| \leqslant 2\delta'$. Clearly, $\lim_{r \to 1-} P_r(t) = 0$ uniformly in t when $\delta' \leqslant |t| \leqslant \pi + \delta'$. Thus there is a positive $\rho < 1$ such that

$$\int_{2\delta' \leqslant |t - \theta_0| \leqslant \pi} P_r(\theta - t)\, dt < \frac{\varepsilon}{8a} \tag{2}$$

whenever $\rho < r < 1$ and $|\theta - \theta_0| \leqslant \delta'$. In view of (1) and (2), for these r and θ we have

$$\begin{aligned} |u_r(\theta) - f(\theta_0)| &\leqslant \int_{\theta_0 - \pi}^{\theta_0 + \pi} P_r(\theta - t)\, |f(t) - f(\theta_0)|\, dt \\ &\leqslant 2a \int_{2\delta' \leqslant |t - \theta_0| \leqslant \pi} P_r(\theta - t)\, dt + \frac{\varepsilon}{4} \int_{|t - \theta_0| \leqslant 2\delta'} P_r(\theta - t)\, dt \\ &< \frac{\varepsilon}{2} \end{aligned}$$

and consequently, $|u_r(\theta) - u_r(\theta_0)| < \varepsilon$. Since u is uniformly continuous in $\{z \in \mathbf{C}\colon |z| \leqslant \rho\}$, there is a $\delta'' > 0$ such that $|u_r(\theta) - u_r(\theta_0)| < \varepsilon$ whenever $0 \leqslant r \leqslant \rho$ and $|\theta - \theta_0| < \delta''$. Now letting $\delta = \min\{\delta', \delta''\}$, we see that the family $\{u_r\colon 0 \leqslant r < 1\}$ is equicontinuous at θ_0.

Conversely, suppose that the family $\{u_r\colon 0 \leqslant r < 1\}$ is equicontinuous at each $\theta \in E$, and choose a sequence $\{r_n\}$ in $[0, 1)$ converging to 1. By the Ascoli theorem (see [**4**; Chapter 9, Corollary 34, p. 179]), the sequence $\{u_{r_n}\}$ contains a convergent subsequence, still denoted by $\{u_{r_n}\}$, which converges at each $\theta \in E$. We set

$$f(t) = \liminf_{n \to \infty} u_{r_n}(t)$$

for each $t \in \mathbf{R}$. Then f is bounded and measurable (see [**4**; Theorem 20, p. 67]), and the equicontinuity of $\{u_{r_n}\}$ at each $\theta \in E$ implies the continuity of f at each $\theta \in E$. Thus $f \in \mathscr{C}_E$, and it suffices to show that $u = P * f$. To this end, observe that for a fixed integer $n \geqslant 1$, the function $z \mapsto u(r_n z)$ is harmonic in D and

continuous in $D \cup T$. By [**6**, Theorem 11.9, p. 235],

$$u(r_n re^{i\theta}) = \int_{-\pi}^{\pi} P_r(\theta - t) u_{r_n}(t)\, dt \tag{3}$$

for each $re^{i\theta}$ in D. Since u is bounded, it follows from [**1**, Corollary 1, p. 5] that $f(t) = \lim_{n\to\infty} u_{r_n}(t)$ for almost all $t \in \mathbf{R}$. Thus fixing r and θ, we have

$$\lim_{n\to\infty} P_r(\theta - t) u_{r_n}(t) = P_r(\theta - t) f(t) \tag{4}$$

for almost all $t \in \mathbf{R}$. Moreover, in $[-\pi, \pi]$ the sequence $\{P_r(\theta - t)u_{r_n}(t)\}$ has an integrable majorant $bP_r(\theta - t)$ where $b = \sup\{|u(z)|: z \in D\}$. Consequently, by the continuity of u and the dominated convergence theorem (see [**4**; Theorem 15, p. 88]), equations (3) and (4) imply that

$$u(re^{i\theta}) = \int_{-\pi}^{\pi} P_r(\theta - t) f(t)\, dt$$

and the proposition is proved.

COROLLARY. *Let $u \in h^\infty$, and let $u_r(\theta) = u(re^{i\theta})$ for $0 \leqslant r < 1$ and $\theta \in \mathbf{R}$. There is an $f \in \mathscr{R}$ with $P * f = u$ if and only if there is a set $E \subset (-\pi, \pi)$ of measure 2π and such that the family $\{u_r: 0 \leqslant r < 1\}$ is equicontinuous at each $\theta \in E$.*

Indeed, $\mathscr{R} = \bigcup_E \mathscr{C}_E$ where the union is taken over all sets $E \subset (-\pi, \pi)$ whose measure is 2π.

REFERENCES

1. P. L. Duren, Theory of H^p Spaces, Academic Press, New York, 1970.
2. D. Minda, The Dirichlet problem for a disk, this MONTHLY, 97 (1990) 220–223.
3. I. I. Priwalow, Randeigenschaften Analytischer Funktionen, VEB Deutscher Verlag der Wissenschaften, Berlin, 1956.
4. H. L. Royden, Real Analysis, MacMillan, New York, 1968.
5. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1976.
6. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1987.

Some Remarks on Weak Continuity

JAMES CHEW
Department of Mathematics, North Carolina A & T State University, Greensboro, NC 27411

JINGCHENG TONG
Department of Mathematics, University of North Florida, Jacksonville, FL 32216

1. Weak continuity and connectedness. As a generalization of a continuous function, Levine [**2**] gave the following definition of a weakly continuous function.

Definition 1. A function $f: X \to Y$ is weakly continuous at $x \in X$ if given any open set V in Y containing $f(x)$, there exists an open set U in X containing x such that $f(U) \subset \overline{V}$. If this condition is satisfied at each $x \in X$, then f is said to be weakly continuous.

continuous in $D \cup T$. By [**6**, Theorem 11.9, p. 235],

$$u(r_n re^{i\theta}) = \int_{-\pi}^{\pi} P_r(\theta - t) u_{r_n}(t)\, dt \tag{3}$$

for each $re^{i\theta}$ in D. Since u is bounded, it follows from [**1**, Corollary 1, p. 5] that $f(t) = \lim_{n\to\infty} u_{r_n}(t)$ for almost all $t \in \mathbf{R}$. Thus fixing r and θ, we have

$$\lim_{n\to\infty} P_r(\theta - t) u_{r_n}(t) = P_r(\theta - t) f(t) \tag{4}$$

for almost all $t \in \mathbf{R}$. Moreover, in $[-\pi, \pi]$ the sequence $\{P_r(\theta - t)u_{r_n}(t)\}$ has an integrable majorant $bP_r(\theta - t)$ where $b = \sup\{|u(z)|: z \in D\}$. Consequently, by the continuity of u and the dominated convergence theorem (see [**4**; Theorem 15, p. 88]), equations (3) and (4) imply that

$$u(re^{i\theta}) = \int_{-\pi}^{\pi} P_r(\theta - t) f(t)\, dt$$

and the proposition is proved.

COROLLARY. *Let $u \in h^\infty$, and let $u_r(\theta) = u(re^{i\theta})$ for $0 \leqslant r < 1$ and $\theta \in \mathbf{R}$. There is an $f \in \mathscr{R}$ with $P * f = u$ if and only if there is a set $E \subset (-\pi, \pi)$ of measure 2π and such that the family $\{u_r: 0 \leqslant r < 1\}$ is equicontinuous at each $\theta \in E$.*

Indeed, $\mathscr{R} = \bigcup_E \mathscr{C}_E$ where the union is taken over all sets $E \subset (-\pi, \pi)$ whose measure is 2π.

REFERENCES

1. P. L. Duren, Theory of H^p Spaces, Academic Press, New York, 1970.
2. D. Minda, The Dirichlet problem for a disk, this MONTHLY, 97 (1990) 220–223.
3. I. I. Priwalow, Randeigenschaften Analytischer Funktionen, VEB Deutscher Verlag der Wissenschaften, Berlin, 1956.
4. H. L. Royden, Real Analysis, MacMillan, New York, 1968.
5. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1976.
6. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1987.

Some Remarks on Weak Continuity

JAMES CHEW
Department of Mathematics, North Carolina A & T State University, Greensboro, NC 27411

JINGCHENG TONG
Department of Mathematics, University of North Florida, Jacksonville, FL 32216

1. Weak continuity and connectedness. As a generalization of a continuous function, Levine [**2**] gave the following definition of a weakly continuous function.

Definition 1. A function $f: X \to Y$ is weakly continuous at $x \in X$ if given any open set V in Y containing $f(x)$, there exists an open set U in X containing x such that $f(U) \subset \overline{V}$. If this condition is satisfied at each $x \in X$, then f is said to be weakly continuous.

Another generalization of a continuous function was given in [3] by Andrew and Wittlesy.

Definition 2. A function $f: X \to Y$ is closure continuous at $x \in X$ if given any open set V in Y containing $f(x)$, there exists an open set U in X containing x such that $f(\overline{U}) \subset \overline{V}$. If this condition is satisfied at each $x \in X$, then f is said to be closure continuous.

Obviously closure continuity implies weak continuity. The converse is not true.

Example 1. Let $X = \{x, y, z, w\}$ with topology $\{\varnothing, \{x, y, z\}, \{z\}, \{z, w\}, X\}$. Let $Y = \{a, b, c, d\}$ with topology $\{\varnothing, \{a, b\}, \{b\}, \{d\}, \{b, d\}, \{a, b, d\}, \{b, c, d\}, Y\}$. Define $f(x) = a, f(y) = b, f(z) = c, f(w) = d$. Then f is weakly continuous since $\{\overline{a, b}\} = \{\overline{b}\} = \{a, b, c\}, \{\overline{b, c, d}\} = Y, \{\overline{d}\} = \{c, d\}$. But f is not closure continuous since $\{\overline{x, y, z}\} = X$.

Theorem 1 in [1] states that a closure continuous function $f: X \to Y$ is connected; i.e., if C is a connected set in X then $f(C)$ is a connected set in Y. The editor of this MONTHLY has forwarded to us communication from Professor J. Ferrer, pointing out that this assertion is false. He gave the following counterexample.

Example 2. Let $X = R$, the real line with the usual topology and let $Y =$ the set of non-negative integers with the co-finite topology; i.e., a set is open in Y if it is empty or its complement is finite. Since the closure of each non-empty open set is all of Y, any function $f: X \to Y$ is closure continuous. Now let $f: X \to Y$ be defined by f (rationals) $= 0$; f (irrationals) $= 1$. Then $f(X) = \{0, 1\}$, a disconnected set in Y. Hence, f is not connected.

The space Y in the above example is T_1 but not T_2. One might wonder whether a closure continuous function into a Hausdorff space is connected. The answer is no even if the range space Y is Urysohn, as the following example shows.

Example 3. Let P be the upper half plane and L be the x-axis. Let $X = P \cup L$. If $\mathscr{T}^*$ is the half-disc topology on X, then $(X, \mathscr{T}^*)$ is a Urysohn space (see [4], Example 78, p. 96). Let $\mathscr{T}$ be the relative topology that X interits by virtue of being a subspace of R^2. The identity function $f: (X, \mathscr{T}) \to (X, \mathscr{T}^*)$ is closure continuous but f takes the connected set L in $(X, \mathscr{T})$ onto the discrete set L in $(X, \mathscr{T}^*)$, showing that a closure continuous function can fail to be connected even in the presence of strong separation properties.

Theorem 1 in [1] should be replaced by the theorem below.

THEOREM 1. *Let X be a connected space and let $f: X \to Y$ be a weakly continuous function from X onto Y $(Y = f(X))$. Then Y is connected.*

Proof. Change the statement "$f(\overline{U}) \subset \overline{V}$" in the proof of Theorem 1 in [1] to "$f(U) \subset \overline{V}$".

Actually we have a little stronger conclusion: if $f: X \to Y$ is weakly continuous and C is a connected set in X such that $Y = f(C)$, then Y is connected.

2. Weak continuity and continuity. The following result is Theorem 1 in [2].

THEOREM 2. *The function $f: X \to Y$ is weakly continuous if and only if for each open set V in Y, $f^{-1}(V) \subset \operatorname{Int} f^{-1}(\overline{V})$.*

COROLLARY 1. *If $f: X \to Y$ is weakly continuous, then for each open set V in Y we have $f^{-1}(V) \subset f^{-1}(\overline{V}) \cap W$, where W is an open set in X.*

Proof. Take $W = \operatorname{Int} f^{-1}(\overline{V})$.

Corollary 1 suggests the following generalization of continuity.

Definition 3. A function $f: X \to Y$ is relatively continuous at $x \in X$ if given an open set V in Y containing $f(x)$, the set $f^{-1}(V)$ is an open set in the subspace $f^{-1}(\overline{V})$. If this condition is satisfied for each $x \in X$, then f is said to be relatively continuous.

If a function $f: X \to Y$ is continuous, then for each open set V in Y, $f^{-1}(V)$ is open in X. Since $f^{-1}(V) \subset f^{-1}(\overline{V})$, we know that $f^{-1}(V)$ is an open set in the subspace $f^{-1}(\overline{V})$. Therefore the following theorem has been established.

THEOREM 3. *A continuous function $f: X \to Y$ is relatively continuous. The converse of Theorem* 3 *is not true.*

Example 4. Let $X = R$ be the real line with usual topology. Let $Y = \{a, b\}$ with discrete topology. Define $f(x) = a$ for $x \in (-\infty, 0]$ and $f(x) = b$ for $x \in (0, +\infty)$. Then f is relatively continuous since $f^{-1}(\{a\}) = f^{-1}(\{\overline{a}\}) = (-\infty, 0] = f^{-1}(\{\overline{a}\}) \cap (-\infty, 1)$ and $f^{-1}(\{b\}) = f^{-1}(\{\overline{b}\}) = (0, +\infty)$. It is easily seen that f is not weakly continuous at $x = 0$.

The significance of relative continuity is that it yields a decomposition of continuity with weak continuity as the other factor.

THEOREM 4. *A function $f: X \to Y$ is continuous if and only if it is weakly continuous and relatively continuous.*

Proof. The necessity is trivial. We prove the sufficiency.

Let V be an open set in Y. By relative continuity, $f^{-1}(V) = f^{-1}(\overline{V}) \cap W$, where W is an open set in X. To prove the openness of $f^{-1}(V)$ in X, let $x \in f^{-1}(V)$. Then $f(x) \in V$ and $x \in W$. By weak continuity of f, there exists an open set U in X containing x such that $f(U) \subset \overline{V}$. This means $U \subset f^{-1}(\overline{V})$. Since the open set W contains x, we may assume $U \subset W$. It now follows that $x \in U \subset f^{-1}(\overline{V}) \cap W = f^{-1}(V)$. This shows that each point $x \in f^{-1}(V)$ is an interior point and so $f^{-1}(V)$ is open in X.

Theorem 4 answers a question raised in [**1**].

3. Weak continuity and interiority. In this section we give a sufficient condition for continuity, which involves weak continuity.

Definition 4. We say that a function $f: X \to Y$ satisfies the interiority condition if $\operatorname{Int} f^{-1}(\overline{V}) \subset f^{-1}(V)$ for each open set V in Y.

Continuity does not imply interiority condition. See the example below.

Example 5. Let $X = Y =$ the set of natural numbers with the co-finite topology. The identity function $f: X \to Y$ is surely continuous. But if V is any proper non-empty open set in Y, then $\overline{V} = Y$ and so $\operatorname{Int} f^{-1}(\overline{V}) = X$ while $f^{-1}(V) = V$. Hence $\operatorname{Int} f^{-1}(\overline{V}) \not\subset f^{-1}(V)$.

THEOREM 5. *Weak continuity plus interiority condition implies continuity.*

Proof. Let $f: X \to Y$ be a weakly continuous function satisfying the interiority condition. Let V be a given open set in Y. By weak continuity, we have $f^{-1}(V) \subset \operatorname{Int} f^{-1}(\overline{V})$. By interiority condition, we have $\operatorname{Int} f^{-1}(\overline{V}) \subset f^{-1}(V)$. Hence $f^{-1}(V) = \operatorname{Int} f^{-1}(V)$, it is open in X.

Relative continuity plus interiority condition does not imply continuity. See Example 4.

Interiority condition does not imply either weak continuity or relative continuity. See example below.

Example 5. Let $X = \{0, 1\}$ be endowed with the co-countable topology, i.e., a set is open if it is empty or its complement is countable. Let $Y = \{0^*, 1^*\}$ with discrete topology. Define $f: X \to Y$ by f (rationals) $= 0^*$, f (irrationals) $= 1^*$. Then f satisfies the interiority condition, but f is neither weakly continuous nor relatively continuous.

4. Between weak continuity and continuity. We conclude with two results which indicate the extent to which weak continuity approximates continuity.

THEOREM 6. *Let $f: X \to Y$ be a weakly continuous function from a space X to a Hausdorff space Y. Then f has closed point inverses.*

Proof. Let $q \in Y$. We must show that $f^{-1}(q) = \{x \in X: f(x) = q\}$ is closed in X, or equivalently $A = \{x \in X: f(x) \neq q\}$ is open in X. To this end, let $x \in A$. Since $f(x) \neq q$ and Y is Hausdorff, there exists an open set V in Y containing $f(x)$ such that $q \notin \overline{V}$. By weak continuity of f, there exists an open set U in X containing x such that $f(U) \subset \overline{V}$. We can prove $U \subset A$. If $U \not\subset A$, there exists $x' \in U$ such that $f(x') = q$. Since $f(U) \subset \overline{V}$, it follows $f(x') = q \in \overline{V}$, contradicting the choice of V.

Example 2 shows that Hausdorffness can not be replaced by T_1 in Theorem 6.

Recall that a subset of a topological space is called closure compact if each open cover of the set contains a finite subcollection whose closures cover the set.

THEOREM 7. *Let $f: X \to Y$ be weakly continuous and let K be a compact set in X. Then $f(K)$ is a closure compact subset of Y.*

Proof. Let $\mathscr{V}$ be an open cover of $f(K)$ and set $\mathscr{V}' = \{V \in \mathscr{V}: V \cap f(K) \neq \emptyset\}$. Then $\mathscr{V}'$ is an open cover of $f(K)$. For each $k \in K$, $f(k) \in V_k$ for some $V_k \in \mathscr{V}'$. By weak continuity of f, there exists an open set U_k in X containing k such that $f(U_k) \subset \overline{V}_k$. The collection $\{U_k: k \in K\}$ is an open cover of K and so there is a finite subcover $\{U_k: k \in K_0\}$, where K_0 is a finite subset of K. Clearly $\{\overline{V}_k: k \in K_0\}$ covers $f(K)$, or $f(K)$ is a closure compact subset in Y.

Added in the Proof. Let $X = \mathbb{R}^2 - \{\text{origin}\}$ with the usual topology and $Y = \mathbb{R}^2$ with the usual topology. It is well known that the familiar function (from complex variables) $f(z) = \ln|z| + i \operatorname{Arg} z$ where $-\pi < \operatorname{Arg} z \leq \pi$ fails to be continuous on the negative reals. We leave it to the reader to verify that f is not even weakly continuous. However, f does satisfy the inferiority condition and f is relatively continuous.

REFERENCES

1. J. Chew, A note on closure continuity, this MONTHLY, 95 (1988) 744–745.
2. N. Levine, A decomposition of continuity in topological spaces, this MONTHLY, 68 (1961) 44–46.
3. D. R. Andrew and E. K. Whittlesy, Closure continuity, this MONTHLY, 73 (1966) 758–759.
4. L. Steen and J. A. Seebach, Jr., Counterexamples in Topology, Holt, Rinehart and Winston, Inc., New York, 1970.

On the Product of Two Alternating Matrices

DRAGOMIR Ž. ĐOKOVIĆ
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

We consider n by n matrices $A = (a_{ij})$ over a commutative ring R. By A' we denote the transpose of A. We say that A is *alternating* if $A' = -A$ and $a_{ii} = 0$ for $i = 1, \ldots, n$. Our main objective is to prove the theorem stated below. The article may be also of interest to the readers who are not familiar with the purely algebraic method of specialization which can often be used as a substitute for the usual topological arguments in the setting of finite dimensional vector spaces over **R** or **C**.

THEOREM. *Let $A = BC$ where B and C are alternating n by n matrices over R, let $f(\lambda)$ be the characteristic polynomial of A, and assume that n is even, say $n = 2m$. Then there exists a monic polynomial $p(\lambda)$ over R of degree m such that*

$$f(\lambda) = p(\lambda)^2, \qquad p(A) = 0. \tag{1}$$

REMARK. When R is a field of characteristic $\neq 2$ it was shown in an Elementary Problem in [**2**] that A has no simple eigenvalues. In fact the solution of that problem given by O. P. Lossers shows that the geometric multiplicity of each eigenvalue of A is at least 2. Both of these results are obvious consequences of our theorem.

We recall some properties of pfaffians and refer the reader to Lang's book [**1**] for more details. Let X be an alternating n by n matrix over R. If n is odd then $\det(X) = 0$. If $n = 2m$ is even then $\det(X) = \mathrm{Pf}(X)^2$ where $\mathrm{Pf}(X)$, the pfaffian of X, is a homogeneous polynomial of degree m in the entries of X. Consequently we have $\mathrm{Pf}(X') = \mathrm{Pf}(-X) = (-1)^m \mathrm{Pf}(X)$. If we assume in addition that X is invertible then X^{-1} is also alternating. Indeed $XX^{-1} = I_n$ implies that $(X^{-1})'X' = I_n$, and so $(X^{-1})' = -X^{-1}$. Furthermore all diagonal entries of X^{-1} are zero since the cofactors of X corresponding to diagonal entries are zero.

LEMMA. *If X is an invertible alternating n by n matrix over R, $n = 2m$, then*

$$\mathrm{Pf}(X^{-1}) \cdot \mathrm{Pf}(X) = (-1)^m.$$

Proof. If Y and Z are n by n matrices over R and Y is alternating, then

$$\mathrm{Pf}(ZYZ') = \det(Z)\mathrm{Pf}(Y),$$

see [**1**, XIV, §10, Theorem 7]. By setting $Y = X^{-1}$, $Z = X$, we obtain $\mathrm{Pf}(X') = \det(X) \cdot \mathrm{Pf}(X^{-1})$. Since $\det(X) = \mathrm{Pf}(X)^2$, the assertion of the lemma follows. ∎

We are now ready to prove the theorem.

Proof of the Theorem. Let x_{ij} and y_{ij}, $1 \leqslant i < j \leqslant n$, be independent indeterminates over the ring of integers **Z**. Let S denote the corresponding polynomial ring $\mathbf{Z}[x_{ij}, y_{ij}; 1 \leqslant i < j \leqslant n]$. Denote by X resp. Y the n by n alternating matrix over S whose (i, j)-th entry is x_{ij} resp. y_{ij} for $1 \leqslant i < j \leqslant n$. Assume that there exists a monic polynomial $q(\lambda)$ over S of degree m such that $\det(\lambda I_n - XY) = q(\lambda)^2$ and

$q(XY) = 0$. Let $\phi\colon S \to R$ be the homomorphism which sends each $x_{ij} \mapsto a_{ij}$ and each $y_{ij} \mapsto b_{ij}$. Then $\phi(X) = B$, $\phi(Y) = C$ and if $p = \phi(q)$ we see that (1) holds.

Hence it suffices to prove (1) in the special case where $R = S$, $B = X$, and $C = Y$. We may view X and Y as matrices over the field of fractions K of S. Then X is invertible, X^{-1} is alternating, and we have

$$\det(\lambda I_n - XY) = \det(X) \cdot \det(\lambda X^{-1} - Y) = \mathrm{Pf}(X)^2 \cdot \mathrm{Pf}(\lambda X^{-1} - Y)^2.$$

Hence, if $p(\lambda) = (-1)^m \mathrm{Pf}(X) \cdot \mathrm{Pf}(\lambda X^{-1} - Y)$, then $f(\lambda) = p(\lambda)^2$ and, by the lemma, p is a monic polynomial in λ of degree m. Next we need Gauss Lemma and the fact that R is a unique factorization domain, see [**1**, V, §6]. Since $p(\lambda) \in K[\lambda]$ and $f(\lambda) \in R[\lambda]$, the equality $f(\lambda) = p(\lambda)^2$ and the Gauss Lemma imply that $p(\lambda) \in R[\lambda]$. This proves the first assertion of (1).

We claim that all roots of $p(\lambda)$, in some algebraic closure E of K, are simple. For this purpose it suffices to show that some specialization of $p(\lambda)$ has only simple roots. For instance we can take the specialization $\phi\colon R \to \mathbf{Z}$ such that

$$\phi(X) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\phi(Y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & m \\ -m & 0 \end{pmatrix}.$$

Then $p(\lambda) \mapsto (\lambda + 1)(\lambda + 2) \cdots (\lambda + m)$ and our claim is established.

At this stage of the proof we know that every eigenvalue, say $\mu \in E$, of $A = XY$ has algebraic multiplicity 2 and so

$$1 \leqslant \dim \ker(\mu I_n - A) \leqslant 2.$$

Since $\mu I_n - A = X(\mu X^{-1} - Y)$ and $\mu X^{-1} - Y$ is alternating, we conclude that $\ker(\mu I_n - A) = \ker(\mu X^{-1} - Y)$ and that $\ker(\mu I_n - A)$ has even dimension. Consequently the geometric multiplicity of μ is 2 and so $p(A) = 0$. Thus the second assertion of (1) also holds. ■

The reader may now wonder what can be said about the case when n is odd. The answer can be obtained easily from the theorem.

COROLLARY. *Let A, B, C and $f(\lambda)$ be as in the theorem except that now n is odd, say $n = 2m + 1$. Then there exists a monic polynomial $p(\lambda)$ over R of degree m such that*

$$f(\lambda) = \lambda p(\lambda)^2, \qquad Ap(A) = 0. \tag{2}$$

Proof. Let $B_1 = B \oplus (0)$, $C_1 = C \oplus (0)$, and $A_1 = B_1 C_1$. By the Theorem we have

$$\det(\lambda I_{n+1} - A_1) = p_1(\lambda)^2$$

and $p_1(A_1) = 0$ where $p_1(\lambda)$ is a monic polynomial over R of degree $m + 1$. Since the last row of A_1 is zero, we have $\lambda | p_1(\lambda)^2$ and consequently $p_1(\lambda) = \lambda p(\lambda)$ and $f(\lambda) = \lambda p(\lambda)^2$. From $p_1(A_1) = 0$ it follows that $p_1(A) = 0$. Thus (2) holds. ■

REFERENCES

1. S. Lang, *Algebra*, Addison-Wesley, Reading 1965.
2. M. Lutzky, Elementary Problem E3189, *Amer. Math. Monthly*, 94 (1987) 181 and 96 (1989) 259–260.

Reliability, Recursion, and Risk

LAVON B. PAGE
Department of Mathematics, North Carolina State University, Raleigh NC 27695

JO ELLEN PERRY
Department of Computer Science, North Carolina State University, Raleigh NC 27695

Two of the most useful elementary facts about probability are these:

$$P(A \cap B) = P(A)P(B) \tag{1}$$

if A and B are independent events;

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$

always. Numerous exercises based on these ideas are found in every introductory text. What is not widely recognized, however, is that one can get a lot of mileage from these (or related) elementary equations even in complex situations if they are used recursively in a proper manner. Recursion and a few elementary facts about probability can provide a basis for calculating the risks inherent in very complicated structures.

The need to assess the risks found in complex technological ventures has become readily apparent with disasters such as Three Mile Island, the Union Carbide accident at Bhopal, India, the space shuttle Challenger's explosion in 1986, and most recently the Exxon Oil spill at Valdez. Engineers have long wrestled with such reliability problems using techniques quite different from the ones we will present here. Our discussion will lead to some contemporary topics in discrete mathematics such as trees and computational complexity and toward implementation of a simple reliability algorithm in Pascal, illustrating in the process the advantage that Pascal has over FORTRAN as a language for discrete mathematics. A brief concluding commentary will indicate the impact that the ideas described here have had on reliability research during the past decade.

FIGURE 1 provides an illustration of a type of model widely used. This figure is a simple example of a *fault tree*. The *leaf nodes* are nodes 5, 7, 8, 9, 10, and 11, and the other nodes are *internal nodes*. The tree is a *full binary* tree in that each node has either two children or no children. The components of the fault tree are of three types: AND gates (intersections), OR gates (unions), and basic events. (The basic events correspond to the leaf nodes.) The purpose of the fault tree is to represent the top event (labeled "system failure" in this case) in terms of unions and intersections of basic events, which are assumed to be mutually independent and whose probabilities of occurrence are assumed to be known. In real applications (such as nuclear or chemical reactors or the space shuttle), fault trees are used to model failure modes of very complex systems, and in such cases evaluation of top-event probability is an important measure of the reliability of the system. In the case of FIGURE 1, system failure occurs if there is both a power outage and

cooling system failure. A power outage may be caused by either sabotage or by a transformer failure, whereas the cooling system fails if a valve sticks or the pump fails. Transformer failure and pump failure are in turn defined as unions or intersections of other basic events. For simplicity we shall assume throughout this discussion that fault trees are represented in binary form. This involves no loss of generality since unions or intersections of more that two events may be rewritten in terms of binary operations simply by introducing parentheses. (In other words, a non-binary fault tree can always be redrawn as a binary tree by introducing additional nodes.)

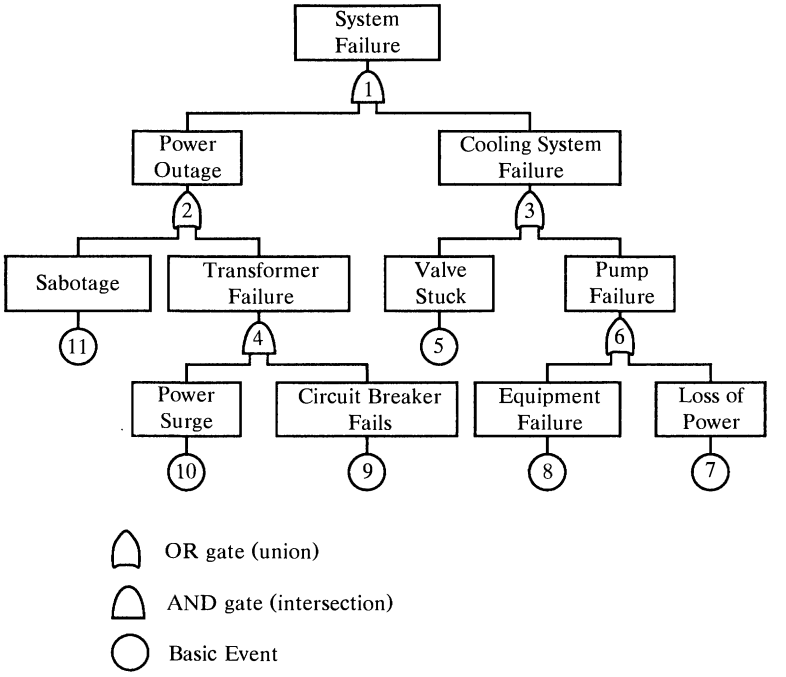


FIG. 1. A simple fault tree.

If it is known that the leaf nodes in the fault tree are mutually independent, then an easy bottom-up hand calculation suffices to calculate the probability of the top event. Starting at the bottom one works up the tree calculating the probability of every event encountered using equations (1) and (2).

For reasons that will become apparent shortly, however, it's much preferable to take a top-down point of view and to think about the problem in terms of recursion. To do so let's first denote by p_n the probability that the event corresponding to node n in the fault tree does in fact occur. Then the goal is to evaluate p_1, and a recursive top-down calculation can proceed as follows:

$$p_1 = p_2 p_3 = (p_4 + p_{11} - p_4 p_{11})(p_5 + p_6 - p_5 p_6).$$

The computation is finished when p_4 and p_6 are replaced by

$$p_4 = p_9 p_{10}$$
$$p_6 = p_7 + p_8 - p_7 p_8.$$

If instead of writing p_n we write $p(n)$, then we can view the above calculation as evaluation of a single value of the recursive function p defined by:

(*i*) $p(n)$ = probability of node n if this node is a basic event,

(*ii*) $p(n) = p(j)p(k)$ if n is an AND gate with children j and k,

(*iii*) $p(n) = p(j) + p(k) - p(j)p(k)$ if n is an OR gate with children j and k.

To eliminate the duplication of function evaluations in (iii), it is preferable in this case to write

(*iii*) $p(n) = 1 - (1 - p(j))(1 - p(k))$.

If this were the end of the story, then fault trees would be a trivial concept. By the procedure just described, the probability of the top event could be determined by calculating a single value of a recursive function, and the nature of the recursion requires that one recursive function call be made for each node in the fault tree. This fact is implicit in (i)–(iii) if one removes the duplication in (iii) as suggested. Therefore the amount of computation required to determine the probability of the top event is directly proportional to the number of nodes in the tree. This is a "linear time" algorithm and represents a happy state of affairs. It's equally easy to observe that the bottom-up approach also requires visiting each node in the tree only once, so in terms of the amount of calculation required the two approaches are equivalent.

What makes the real world infinitely more interesting (and complex) than this simple example is the fact that events often are *not* independent. Consider FIGURE 2 for example. The only feature that is changed from FIGURE 1 is found in the contributing factors to POWER OUTAGE, the event labeled node 2 in the fault tree. SABOTAGE has been eliminated as a cause, and LOSS OF POWER now appears. This basic event LOSS POWER is the same event that leads to PUMP FAILURE, event number 6. So now the mutual independence of different parts of the tree is destroyed. (Notice that the two "loss of power" events are now given the same label "7." This is necessary since basic events with distinct labels are assumed to be mutually independent.) The number of *distinct* basic events is reduced from six to five, but even if we consider these five still to be mutually independent, the fact that one of them appears at two different locations in the tree means that neither the bottom-up nor the top-down procedure previously used will now be valid. The reason is that both procedures assume that whenever an AND or OR gate appears in the tree, the children will represent independent events. This is no longer the case for the top node in the tree. Since LOSS OF POWER is a contributing factor to both event 2 and event 3, the children of event 1 are no longer independent.

We mention at this point that duplication of interior node labels at multiple locations in the fault tree is perfectly acceptable as long as the entire subtrees rooted at each location are identical. For reasons of convenience in the complexity

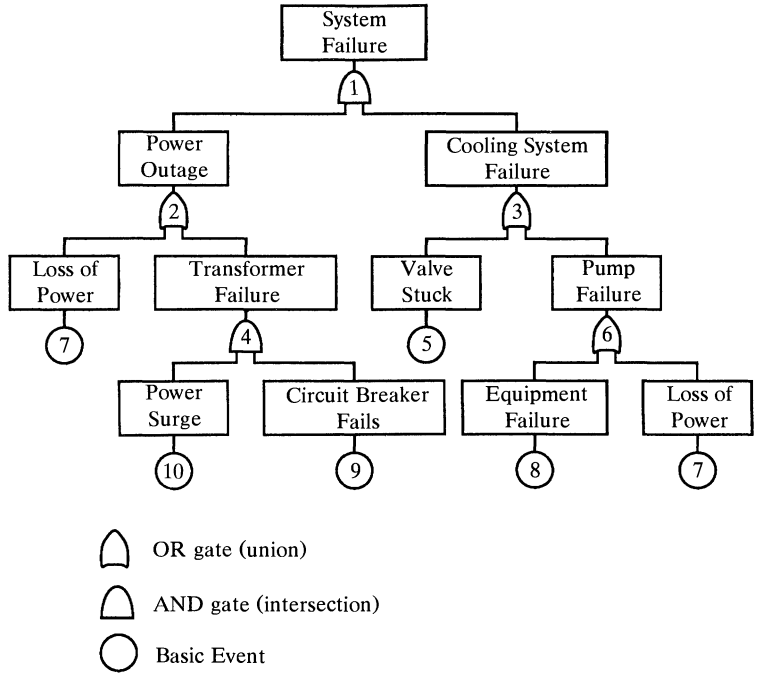


FIG. 2. Appearance of node 7 at two locations in the tree means that AND gates and OR gates no longer always have independent children.

analysis given later, however, we will assume that distinct interior nodes have distinct integer labels.

The net effect of duplicated leaves in the fault tree is to render the bottom-up approach useless and to force reconsideration of the top-down approach. Equation (1) no longer offers a valid means for treating AND gates. Equation (2) is still valid since it doesn't require independence. So OR gates can be treated in terms of equation (2) if we can decide how to handle AND gates. To better understand the situation, we introduce some appropriate notation:

Notation. E_i denotes the event corresponding to the node labeled "i" in the fault tree. Is S is a set of node labels in the fault tree, then the probability that all of the events E_i, $i \in S$, occur is denoted by

$$p_S = P\left(\bigcap_{i \in S} E_i\right).$$

For an AND gate n with dependent children j and k, it is no longer true as in equation (1) that the probability of the AND gate is the product of the probabilities of its children. Using the notation just introduced, however, it is true that $p_{\{n\}} = p_{\{j,k\}}$. More generally, suppose now that S is a set of node labels, that $n \in S$ is the label of an AND gate in the tree, and that j and k are the children of node n. If we let C denote the intersection of the collection of events $\{E_i\colon i \in S$ and $i \neq n\}$, then we have available to us the following recursive mechanism for

eliminating AND gates from the set S:

$$p_S = P\Big(\bigcap_{i\in S} E_i\Big) = P(E_n \cap C) = P(E_j \cap E_k \cap C) = p_{S\cup\{j,k\}-\{n\}}. \tag{3}$$

If n were an OR gate rather than an AND gate and if C denotes the same set as above, then the manner in which we work our way down the tree is a bit different:

$$\begin{aligned} p_S &= P\Big(\bigcap_{i\in S} E_i\Big) = P(E_n \cap C) = P\big[(E_j \cup E_k) \cap C\big] \\ &= P\big[(E_j \cap C) \cup (E_k \cap C)\big] \\ &= P(E_j \cap C) + P(E_k \cap C) - P(E_j \cap E_k \cap C) \\ &= p_{S\cup\{j\}-\{n\}} + p_{S\cup\{k\}-\{n\}} - p_{S\cup\{j,k\}-\{n\}}. \end{aligned} \tag{4}$$

The value of equations (3) and (4) is that they do not require that the events E_j and E_k be independent. This enables us to proceed recursively all the way down the tree from the top event to the basic events without concern as to whether the basic events occur at multiple locations in the tree. When we have finally eliminated all AND gates and OR gates in this manner, we are left with a large number of terms of the form p_T where T is a set of node labels representing basic events in the tree. By assumption these represent mutually independent events, so each such term p_T may be immediately evaluated as

$$p_T = \prod_{k\in T} P(E_k). \tag{5}$$

As an illustration, calculation of the probability of system failure in FIGURE 2 might begin as follows:

$$\begin{aligned} p_{\{1\}} &= p_{\{2,3\}} \\ &= p_{\{3,4\}} + p_{\{3,7\}} - p_{\{3,4,7\}} \\ &= p_{\{3,9,10\}} + (p_{\{5,7\}} + p_{\{6,7\}} - p_{\{5,6,7\}}) - p_{\{3,7,9,10\}}. \end{aligned}$$

In this sample computation, node 1 is first replaced by its children using equation (3). Then node 2 is removed using equation (4). Next the term $p_{\{3,4\}}$ is replaced by $p_{\{3,9,10\}}$ using equation (3), $p_{\{3,7\}}$ is replaced by $p_{\{5,7\}} + p_{\{6,7\}} - p_{\{5,6,7\}}$ by using equation (4), and $p_{\{3,4,7\}}$ becomes $p_{\{3,7,9,10\}}$ by again using equation (3). In the term $p_{\{5,7\}}$ in the last expression, each element of the set $\{5,7\}$ is a label corresponding to a basic event. Therefore $p_{\{5,7\}}$ may be evaluated as the product of the probability of E_5 times the probability of E_7. The remaining terms of the last expression require further replacement of AND or OR gates by their children using equations (3) and (4). The entire computation is mapped in FIGURE 3. The arithmetic itself is not shown in the figure, and only the subscript set S is shown for each term p_S encountered during the computation. The purpose is to show the manner in which each term is evaluated by reference to other terms until finally all AND and OR gates have been removed from every term and equation (5) can be used.

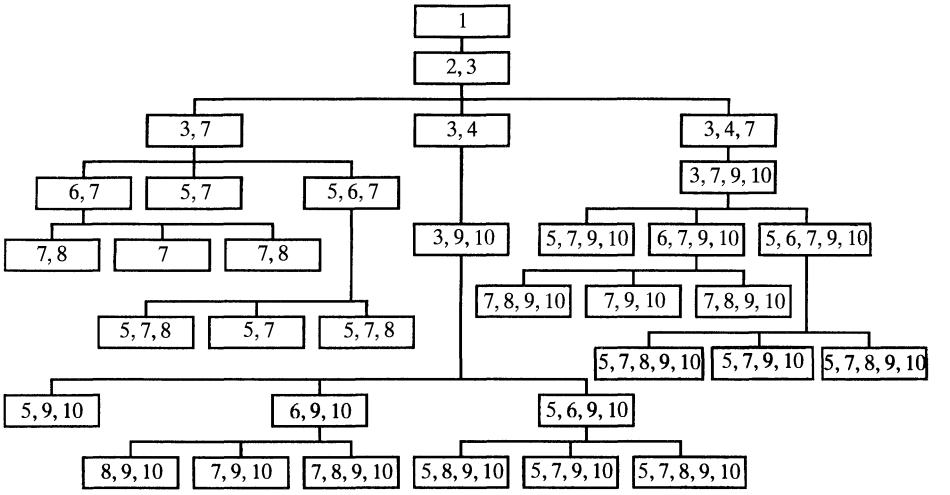


FIG. 3. One possible recursion tree generated by recursively applying equations (3) and (4) to calculate the probability of the top event in the fault tree of FIGURE 2.

FIGURE 3 is a *recursion tree*. While the fault tree itself is binary, each node in the recursion tree is either a leaf, a node with one child, or a node with three children. These cases correspond respectively to these cases in which:

- all nodes in the set are basic events;
- an AND gate is used for pivoting via equation (3);
- an OR gate is used for pivoting via equation (4).

The recursion tree will not be uniquely determined unless one specifies a rule for choosing the precise node to use in the decomposition process at each step.

The shape of the recursion tree is not as important as its size, however, for the size of this tree reflects the amount of computation that must be performed to evaluate the probability of the top event. One measure of size is simply to count the number of nodes in the recursion tree. A more convenient measure, however, is to count only the leaf nodes. One reason that it is preferable to count only the leaves is that this eliminates the quandary over how to terminate the recursion. For example, one could terminate the recursion by expanding the product in equation (5) whenever a set of basic events is encountered. Alternatively, one can continue to think recursively and when S is a set of basic events, compute p_S as $p_S = p_{\{n\}}p_{S-\{n\}}$ where $n \in S$. In the latter case we continue the recursion until S is empty, in which case $p_S = 1$. If we showed this additional recursion in the recursion tree of FIGURE 3, each leaf in the tree as shown would have a "branch" descending down from it in which the basic events are themselves eliminated one-at-a-time until only the empty set remains. Notice that this causes no change in the number of leaves in the recursion tree. So whether this kind of "tail recursion" is used or not will be irrelevant if we measure the complexity of the computation by the number of leaves. Another benefit to counting only the leaves is that this measure of computational complexity is independent of the manner in

which the nodes are selected for "pivoting" via equations (3) and (4) throughout the recursive process.

THEOREM. *Suppose equations* (3) *and* (4) *are used recursively to determine top-event probability for a full binary fault tree in which only basic events appear at multiple locations in the fault tree. Let* ψ *be the recursive function defined on the set of node labels in the fault tree as follows*:

(1) $\psi(n) = 1$ *if* n *is a basic event (leaf node),*

(2) $\psi(n) = \psi(j)\psi(k)$ *if* n *is an AND gate with children* j *and* k,

(3) $\psi(n) = \psi(j) + \psi(k) + \psi(j)\psi(k)$ *if* n *is an OR gate with children* j *and* k.

Then for every node n *in the fault tree,* $\psi(n)$ *gives the number of leaves in the recursion tree produced by using equations* (3) *and* (4) *recursively to evaluate* $P(E_n) = p_{\{n\}}$. (*In particular the number of leaves in the recursion tree generated by calculating the probability of the top event does not depend on the order in which* AND *gates or* OR *gates are selected for decomposition via equations* (3) *and* (4).)

Proof. If S is a set of nodes in the fault tree and m and n are interior nodes (not leaves) in S, it is easy to check that the terms generated by elimination of m and n from S using equations (3) and (4) do not depend on which node (m or n) is eliminated first. This is true whether both are AND nodes, both are OR nodes, or there is one of each. From this observation it follows that the final collection of terms that remain after all interior nodes in S are eliminated in this manner (corresponding to the leaves in the recursion tree) does not depend on the order of application of equations (3) and (4) or the choices of nodes to which these equations are applied.

As for the function ψ, first note that if n is a basic event in the fault tree, then $p_{\{n\}}$ is known and recursion terminates immediately (1 leaf node). So the number of leaves generated in evaluating $p_{\{n\}}$ is 1, which is $\psi(n)$ by definition.

Suppose now that n is an AND gate with children j and k. Suppose also that it is known that the desired property holds for nodes j and k, that is $\psi(j)$ and $\psi(k)$ represent the number of leaves in the recursion tree generated by evaluation of $p_{\{j\}}$ and $p_{\{k\}}$ respectively. The evaluation of $p_{\{n\}}$ begins with the fact that $p_{\{n\}} = p_{\{j,k\}}$. Furthermore, if node k in this set is ignored while j and all its descendents that are not basic events are eliminated via use of equations (3) and (4), the number of terms generated is then $\psi(j)$. (The process is the same as it would be if k were not in the set. If k were not present, then each term so generated would be a leaf in the recursion tree generated by evaluation of $p_{\{j\}}$.) Since k is present, however, the remaining terms will all now consist of k together with some set of basic event node labels. (This is the point in the argument at which the fact that interior nodes have distinct labels is important. If this were not the case, k might have been eliminated from some of the terms in the process of eliminating j and its descendents.) In each remaining term k and all its non-basic-event descendents must be removed, and the manner for doing so is identical to what it would be if k were now the only element in each such set. Therefore each term obtained in removing j and its non-basic descendents leads in turn to $\psi(k)$ leaves in the

```
function P(S)
    input: A set S of positive integers corresponding to node lables in the fault tree
   output: The probability that all the events corresponding to node labels in S do occur
begin
if S is empty then
    P ← 1
else if there exists an AND gate n in S with children j and k then
    P ← P(S ∪ {j, k} − {n})
else if there exists an OR gate n in S with children j and k then
    P ← P(S ∪ {j} − {n}) + P(S ∪ {k} − {n}) − P(S ∪ {j, k} − {n})
else let n denote any basic event label in S then
    P ← P(S − {n}) × probability of event n
end
```

FIG. 4. Pseudo-code for a recursive function to incorporate use of equations (3) and (4). The probability of the top event of a full binary fault tree is obtained by evaluating $P(S)$ where S contains the single node number corresponding to the top event. This pseudo-code is easily translated into Pascal since Pascal supports both set operations and recursion. The requirement that the tree be full binary is not essential, but it simplifies the description of the manner in which AND gates and OR gates are treated.

recursion tree, for a total of $\psi(j)\psi(k) = \psi(n)$ such nodes. Therefore $\psi(n)$ has the desired characterization.

The argument for OR gates is similar. One then starts instead with the equation $p_{\{n\}} = p_{\{j\}} + p_{\{k\}} - p_{\{j,k\}}$. If one assumes that the children of n have the desired property, then $\psi(j)$ is the number of leaves generated in evaluation of $p_{\{j\}}$, $\psi(k)$ is the number of leaves generated in evaluation of $p_{\{k\}}$, and as argued in the preceding paragraph, $\psi(j)\psi(k)$ is the number of leaves generated in evaluation of $p_{\{j,k\}}$. Thus evaluation of $p_{\{n\}}$ leads to $\psi(j) + \psi(k) + \psi(j)\psi(k) = \psi(n)$ leaves.

Because the number of terms generated by use of equations (3) and (4) is so great, hand calculations are not feasible even for relatively small trees. The beauty of equations (3) and (4) is that they represent an easily programmed recursive approach, especially if the programming language supports both recursion and standard set notation (as does Pascal). In fact, the entire manner for evaluating the probability of the top event based on repeated use of equations (3) and (4) is wrapped up in the recursive function definition shown in FIGURE 4. To implement this function in Pascal requires only a few appropriate type declarations to define the data structures used to represent nodes in the fault tree and a simple utility function to return the node number of an AND gate, OR gate, or basic event in S. To evaluate the probability of the top event, one simply issues the initial call to evaluate $P(S)$ where S is the set containing the single node number corresponding to the top event. A Pascal program to evaluate top-event probability based on this recursive function definition consists of fewer than 100 lines of source code, and most of that is used to read the fault tree from an input file. (An easy way to represent the fault tree is to let each line of the input file represent a node in the fault tree and consist of a character indicating node type followed by left and right children in the case of an AND or OR gate or the probability in the case of a basic event.)

The algorithm based on equations (3) and (4) appears in [3]. Its role in fault-tree research is somewhat akin to Euler's method for numerically solving differential

equations. Because of its intuitive appeal and simplicity, its value is primarily pedagogical. Euler's method is shunned for serious work because other methods are much more accurate. In the case of the fault tree algorithm, the limiting factor is the computational complexity as described by the recursive function ψ of the theorem. An attractive feature of the function ψ is that evaluation of ψ requires only a number of computations proportional to the number of nodes in the fault tree. This means that we can use the recursive function ψ as a means of pre-calculating the number of steps required to apply equations (3) and (4) as a means of evaluating top-event probability in any fault tree.

An Illustration. FIGURE 5 has appeared frequently in the scientific literature [1, 2, 4, 5]. Once it was in fact used to demonstrate a special electronic apparatus called ESCAF developed under the auspices of the French Atomic Energy Agency [2]. Notice that there are numerous replicated basic events making large modules in the fault tree dependent upon one another.

The top event is node 21, and this tree is still small enough that a manual calculation of $\psi(21)$ proceeds quickly. Alternatively, it is a simple matter to embed the recursive function ψ in a simple program and to execute the program on the fault tree of FIGURE 5 and learn that $\psi(21) = 1{,}000{,}447$. Therefore use of equations (3) and (4) to evaluate the probability of the top event would generate a recursion tree with 1,000,447 leaves.

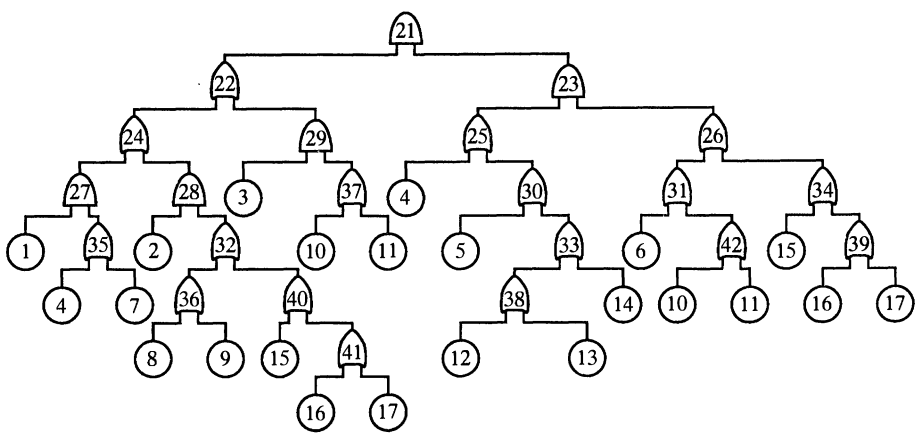


FIG. 5. **Fault tree from [1] which has been used to demonstrate a number of algorithms.**

How big is big? Fault trees of much greater size and complexity than the one in FIGURE 5 are routinely encountered in real applications. Is there no hope for using recursive methods on such trees? Fortunately there is, but the moral of the story is that independence in the fault tree is ignored only at great peril. Whereas the first simple approach examined requires that independence be present everywhere in the fault tree (that is, no duplication of basic events is permitted), the approach based on equations (3) and (4) makes no use of independence whatsoever except in

An "Oceans of Zeros" Proof That a Certain Non-Liouville Number is Transcendental

M. J. KNIGHT
P. O. Box 732, Lompoc, CA 93438

This note contains a simple new proof of the transcendence of the number F:

$$F = \sum_{n=0}^{\infty} (1/B)^{2^n}, \tag{1}$$

for any positive integer $B > 1$, and is well suited to a classroom presentation. It is an argument using only the definition of a transcendental number and the representations of numbers in the bases 2 and B. The powers of F, when written in base B, become more and more difficult to describe, but may be described well enough to yield transcendence.

The first numbers proven transcendental were the Liouville numbers, characterized by having orders of approximation of arbitrarily large sizes (see [3] for background material and additional references on transcendental numbers). The proof that they are transcendental uses Liouville's theorem that the order of approximation possible for an algebraic number of degree m is less than $m + 1$. Although Baker calls the Liouville theorem the first practical criterion for proving transcendence [1], the argument given here could have been used by Liouville to prove that F and many other numbers are transcendental.

The number F looks like it might be a Liouville number, since taking partial sums in (1) leads to an infinite number of very good rational approximations to F. In fact, these give approximations of order 2, not nearly good enough to prove F a Liouville number, nor even good enough to satisfy the very strong transcendental result of Roth [4]. The number F is not a Liouville number, because the continued fraction for F is known [6] and has bounded partial quotients, while the continued fraction for any Liouville number has unbounded partial quotients. The number F is an example of a Fredholm number; such numbers have been shown to be transcendental by different methods [2, 5].

The proof is by contradiction, so assume that F is algebraic and satisfies the polynomial

$$c_e F^e + c_{e-1} F^{e-1} + \cdots + c_0 = 0$$

with $c_i \in \mathbb{Z}$, and $c_e > 0$. Let H be the maximum of the absolute values of the coefficients of the polynomial. Rearranging, we have

$$c_e F^e + \cdots = b_s F^s + \cdots \tag{2}$$

with all coefficients on both sides nonnegative. Think of this as an equation between two base-B representations, with the required carries not yet performed.

The basic trick used is that, in base B, each of the powers of F with exponents less that e have a common, arbitrarily long, span of all zeros, i.e., an "ocean" of zeros, whereas the same span for the eth power has an isolated nonzero "island." By picking such a span long enough, based on e and H, an isolated nonzero island will remain in the expression on the left of equation (2), but the same span will be all zeros on the right. Thus, the equality cannot hold.

We now specify the location of the ocean of zeros and the nonzero island (consideration of the effect of carries will follow).

Let

$$f(x) = \sum_{n=0}^{\infty} x^{2^n},$$

so that $F = f(1/B)$, and let $a(r, k)$ denote the coefficient of x^r in $f(x)^k$. Observe that $a(r, k)$ is equal to the number of ways that r can be written as a sum of k powers of 2, where different orderings of the terms are counted as distinct.

LEMMA 1 (Oceans of Zeros and the Island). *Fix positive integers e and m, and consider the exponents k for which* $1 \leqslant k \leqslant e$. *For all r in the span* $N - (2^{m-1} - 1) \leqslant r \leqslant N + (2^m - 1)$, *with* $N = (2^e - 1)2^m$, $a(r, k)$ *satisfies*

$$a(r, k) = \begin{cases} e!, & \text{if } r = N \text{ and } k = e \\ 0, & \text{otherwise.} \end{cases}$$

Proof. The binary expansion of N is $111\ldots111000\ldots000$, with e 1s and m 0s. If r is in the given interval and not equal to N, then r has more than e 1s in its expansion and $a(r, k)$ is 0 by the remark before the lemma. Suppose $r = N$. Then N can be written as a sum of k powers of 2 only if $k = e$, and in that case it can be done in $e!$ ways, since all $e!$ permutations of the distinct powers of 2 have to be counted. This completes the proof.

Consider equation (2) represented in base B, still without carries, and use Lemma 1 to describe the powers of F for the span of digits given by Lemma 1, with the value of m to be chosen later. The left-hand side will have this span be all zeros, except at digit N. The right-hand side will have this span consisting only of zeros. Both sides will have nonzero entries at some digits beyond $N + (2^m - 1)$.

Now consider all possible carries. The term at digit N, on the left-hand side of equation (2), is $c_e e!$, which is independent of m, so for large m all of its carry will remain in the left half of the span, forming the island.

The carries from all locations beyond the span, for either side of equation (2), will not reach position N. This will follow from Lemma 2, which shows that no single power of F will have carries that reach that far, for the sum on either side consists of at most $e + 1$ terms, each with a positive coefficient bounded by H. But since both e and H are independent of m, the effects of multiplying by H and adding no more than $e + 1$ terms may be made insignificant by choosing m large.

LEMMA 2 (carries for the kth power of F). *Let e, m and N be as in Lemma* 1, *with m large, and let* $R = N + 2^m$. *Then*

$$\sum_{r=R}^{\infty} \frac{a(r, k)}{B^r} < \frac{1}{B^{N+1}}.$$

Proof. First observe that $a(r, k)$ is bounded by the product of $k!$ with the number of all possible choices of k powers of 2 that are not greater than r. This product may be bounded by $(1 + \log_2 r)^k k!$, which, for large r, is bounded by r.

Thus the sum in the lemma is bounded by

$$\sum_{r=R}^{\infty} \frac{r}{B^r} = \left(1 - \frac{1}{B}\right)^{-2}\left[\frac{R}{B^R} + \frac{1-R}{B^{R+1}}\right].$$

This equality may be shown by multiplying the series on the left by $(1 - B^{-1})^2$ and telescoping the result. For m large, the expression on the right is less than $B^{-(N+1)}$, as required.

Thus these carries do not affect the island. The carries, in fact, are insignificant (around m places versus 2^m places needed to reach the island). The nonzero entry from $a(N, e)$ will not be affected by other terms on the left-hand side and not matched by terms on the right-hand side. Therefore, (2) cannot hold and F is transcendental.

Acknowledgement. This paper has benefited greatly from the valuable comments of the referees.

REFERENCES

1. A. Baker, Transcendental Number Theory, Cambridge University Press, New York, 1975.
2. J. H. Loxton and A. J. van der Poorten, Transcendental and algebraic independence by a method of Mahler, chapter 15 of A. Baker and D. W. Masser, Transcendence Theory: Advances and Applications, Academic Press, New York, 1977.
3. Ivan Niven, Irrational Numbers, The Carus Mathematical Monographs, no. 11, The Mathematical Association of America, 1967.
4. K. F. Roth, Rational approximations to algebraic numbers, *Mathematika*, 2 (1955) 1–20.
5. Th. Schneider, Einfuhrung in die Transzendenten Zahlen, Springer, Berlin, 1957.
6. J. Shallit, Simple continued fractions for some irrational numbers, *Journal of Number Theory*, 11 (1979) 209–217.

Simple Tests for Classifying Critical Points of Quadratics with Linear Constraints

PAUL BINDING*
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

1. Introduction. Optimization texts (e.g., [**5**, pp. 234–5]) frequently give second-order tests for classifying (perhaps constrained) critical points of general smooth functions in terms of "accessory problems" involving quadratic forms

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \qquad \mathbf{x}^T = [x_1 x_2 \cdots x_n] \tag{1.1}$$

(perhaps subject to linear constraints of the form $C\mathbf{x} = \mathbf{0}$), where A is $n \times n$ symmetric and C is $m \times n$ of rank $m \leqslant n$, all scalars being real. Advanced texts generally provide no *algorithms* for this classification, however, and the standard determinant tests used in most elementary texts (including those on multivariable calculus) are incomplete, inefficient, and difficult to motivate. The pivot algorithms presented below are simple to understand and to implement, and the sign rules are obvious. I believe that they provide a nice application of the Gauss-type procedures taught extensively to undergraduates.

*Research supported by NSERC of Canada.

Thus the sum in the lemma is bounded by

$$\sum_{r=R}^{\infty} \frac{r}{B^r} = \left(1 - \frac{1}{B}\right)^{-2}\left[\frac{R}{B^R} + \frac{1-R}{B^{R+1}}\right].$$

This equality may be shown by multiplying the series on the left by $(1 - B^{-1})^2$ and telescoping the result. For m large, the expression on the right is less than $B^{-(N+1)}$, as required.

Thus these carries do not affect the island. The carries, in fact, are insignificant (around m places versus 2^m places needed to reach the island). The nonzero entry from $a(N, e)$ will not be affected by other terms on the left-hand side and not matched by terms on the right-hand side. Therefore, (2) cannot hold and F is transcendental.

Acknowledgement. This paper has benefited greatly from the valuable comments of the referees.

REFERENCES

1. A. Baker, Transcendental Number Theory, Cambridge University Press, New York, 1975.
2. J. H. Loxton and A. J. van der Poorten, Transcendental and algebraic independence by a method of Mahler, chapter 15 of A. Baker and D. W. Masser, Transcendence Theory: Advances and Applications, Academic Press, New York, 1977.
3. Ivan Niven, Irrational Numbers, The Carus Mathematical Monographs, no. 11, The Mathematical Association of America, 1967.
4. K. F. Roth, Rational approximations to algebraic numbers, *Mathematika*, 2 (1955) 1–20.
5. Th. Schneider, Einfuhrung in die Transzendenten Zahlen, Springer, Berlin, 1957.
6. J. Shallit, Simple continued fractions for some irrational numbers, *Journal of Number Theory*, 11 (1979) 209–217.

Simple Tests for Classifying Critical Points of Quadratics with Linear Constraints

PAUL BINDING*
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

1. Introduction. Optimization texts (e.g., [5, pp. 234–5]) frequently give second-order tests for classifying (perhaps constrained) critical points of general smooth functions in terms of "accessory problems" involving quadratic forms

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \qquad \mathbf{x}^T = [x_1 x_2 \cdots x_n] \tag{1.1}$$

(perhaps subject to linear constraints of the form $C\mathbf{x} = \mathbf{0}$), where A is $n \times n$ symmetric and C is $m \times n$ of rank $m \leqslant n$, all scalars being real. Advanced texts generally provide no *algorithms* for this classification, however, and the standard determinant tests used in most elementary texts (including those on multivariable calculus) are incomplete, inefficient, and difficult to motivate. The pivot algorithms presented below are simple to understand and to implement, and the sign rules are obvious. I believe that they provide a nice application of the Gauss-type procedures taught extensively to undergraduates.

*Research supported by NSERC of Canada.

In sections 2–4 we discuss a complete definiteness test for (1.1) without constraints, for a symmetric matrix A. If A is definite, i.e., $q(\mathbf{x}) = 0$ only if $\mathbf{x} = \mathbf{0}$, then a simple recursive reduction to the form

$$q(\mathbf{x}) = p_1 y_1^2 + \cdots + p_n y_n^2 \tag{1.2}$$

is possible for suitable linear combinations y_j of the x_j avoiding expensive eigenvalue calculations. The p_j (which all have the same sign) may be calculated via nested subdeterminants of A, via successively completing the square, or as the pivots in the Gauss reduction of A to row echelon form. Completing the square is rejected in favor of determinants in [4], presumably because of the symbolic algebra involved. On the other hand, row-reduction of A needs only about $n^3/6$ multiplications to reduce the pivots. The only way determinants can compete is if they are evaluated as a nested set via pivoting, so one might as well use pivots from the outset, and avoid setting up the determinants.

The situation is more clear-cut when some pivots vanish, since the determinant tests then need significant modification. Most authors who discuss the issue require, either implicitly or explicitly, the earlier nested determinants corresponding to *all permutations* of $x_1, \ldots, x_n$; see e.g., [1, 2]. This is computationally prohibitive for large n. There is an analogous pivot test for nonnegative definiteness in [7, p. 257], but it is not clear how it applies when all the a_{ii} are zero, for example when $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

In Section 5 we admit the constraint $C\mathbf{x} = \mathbf{0}$ by applying the results of Sections 3 and 4 to a "reduced" matrix

$$R = N^T A N \tag{1.3}$$

where the columns of N form a basis for the null space of C. Since N may be computed directly from the row echelon form of C, the complete algorithms are easily implemented by beginners. Moreover it is also easy to see why the method works, in contrast with standard determinant methods, which usually involve the "bordered" matrix $B = \begin{bmatrix} 0 & C \\ C^T & A \end{bmatrix}$, and which we review in Section 6. When pivots vanish, expensive variable permutations are unnecessary, and the pivot method then has a considerable advantage, as we illustrate with a numerical example in Section 7.

2. Basic pivoting. We start with some observations on reduction of the matrix A of (1.1) by elementary operations. If $a_{11} \neq 0$ then $n - 1$ elementary row operations convert each of $a_{21}, \ldots, a_{n1}$ to zero, and leave the first row unchanged. These operations correspond to premultiplication by elementary matrices, and taken together they convert A to $L_1 A$, say, where L_1 is lower triangular. (Iterating this procedure, we obtain the familiar LDU decomposition of A; see [7, p. 25] and Section 3.) By symmetry of A, the corresponding $n - 1$ column operations convert $a_{12}, \ldots, a_{1n}$ to zero, so

$$L_1 A L_1^T = \begin{bmatrix} a_{11} & \mathbf{0}^T \\ \mathbf{0} & A_{>1} \end{bmatrix}, \tag{2.1}$$

where $A_{>1}$ is the result of applying the row and column operations to the bottom

right $(n-1)\times(n-1)$ submatrix of A. Combining (2.1) with (1.1) we have

$$q(\mathbf{x}) = a_{11}y_1^2 + \mathbf{x}_{>1}^T A_{>1}\mathbf{x}_{>1}, \tag{2.2}$$

where y_1 is the first component of $L_1^{-T}\mathbf{x}$ and $\mathbf{x}_{>1}^T = [x_2, \ldots, x_n]$.

Remark. The column operations do not affect $A_{>1}$, so we may obtain it by row operations alone. In fact if we first divide row one of A by a_{11} and next create the zeros underneath, then we obtain

$$L_1' A = \begin{bmatrix} 1 & \mathbf{e}^T \\ \mathbf{0} & A_{>1} \end{bmatrix}, \tag{2.3}$$

where L_1' is lower triangular and $y_1 = [1\ \mathbf{e}^T]\mathbf{x}$. Thus all the information in (2.2) may be obtained from $L_1'A$, provided we store the "pivot" $p_1 = a_{11}$.

Iterating the procedure used for (2.1), we obtain

$$L_i A L_i^T = \begin{bmatrix} D_i & 0 \\ 0 & A_{>i} \end{bmatrix} \tag{2.4}$$

and

$$q(\mathbf{x}) = \sum_{k=1}^{i} p_k y_k^2 + \mathbf{x}_{>i}^T A_{>i}\mathbf{x}_{>i}, \qquad 1 \leqslant i < n \tag{2.5}$$

where U_i is upper triangular, $D_i = \operatorname{diag}(p_1, \ldots, p_i)$, the y_k are components of $L_i^{-T}\mathbf{x}$, and $x_{>i}^T = [x_{i+1}, \ldots, x_n]$. This reduction is possible provided each of the "pivots" $p_k \neq 0$, and it is equivalent to "completing the square" in $q(\mathbf{x})$ to isolate the terms in $x_1, \ldots, x_i$. When $i = n$, we obtain (1.2). Alternative arguments may be found in, say, [**4**, §5.11] or [**7**, §6.2].

3. Unconstrained definiteness. As we have seen, calculation of the right side of (2.1) is a little more efficient (and familiar) via (2.3). Moreover (2.1) also shows that $A_{>1}$ is symmetric, and this can be used as a check on the arithmetic. Alternatively we can complete the $A_{>i}$ by symmetry, saving nearly half the work: indeed it can be shown that only $n(n-1)(n+4)/6$ multiplications are needed to reduce A to an upper triangular matrix U with unit main diagonal. With reference to (2.5), the successive pivots are the p_i, and $\mathbf{y} = U\mathbf{x}$, so perhaps the simplest method of computing the row echelon form of A also yields (1.2). Moreover $\mathbf{y} = U\mathbf{x}$ shows that the y_i are linearly independent combinations of the x_j, so q is positive definite (PD) if the p_i are all positive. Further $a_{ii} = q(\mathbf{e}_i)$ where $\mathbf{e}_i$ is the ith column of I. Thus $p_1 = a_{11} > 0$ if q is PD, and q cannot be PD if any $a_{ii} \leqslant 0$. Similar reasoning applies to the other $A_{>i}$, and leads us to the following.

PD Test.

0. set $i = 1$.
1. Test each a_{jj} and stop if $\leqslant 0$: q is not PD. Stop if $i = n$: q is PD.
2. Calculate $A_{>1}$ as in (2.3)—necessarily $a_{11} \neq 0$.
3. Replace A by $A_{>1}$, i by $i+1$, and return to 1.

The top left $i \times i$ determinants α_i of A can be shown to satisfy

$$p_i = \frac{\alpha_i}{\alpha_{i-1}}, \quad \text{where } \alpha_0 = 1, \tag{3.1}$$

p_i as usual being the pivots for A. Thus PD and ND (negative definiteness) can also be tested via the α_i. Note that the pivot sign rule for ND is obvious, but the determinant sign rule depends on (3.1) or an equivalent argument. Computation is also more direct via the p_i, as we discussed earlier.

4. A general definiteness test. Recall that (2.4) depends on the pivots p_i not vanishing. If some $p_i = 0$ then the first row of $A_{>i}$ either vanishes or else has a nonzero element. In the former case, (2.4) is independent of x_{i+1} and we may set $p_i = 0$ and continue with $A_{>(i+1)}$, which is defined by suppressing row and column one of $A_{>i}$. In the latter case, we may write

$$(A_{>i})_{11} = 0 \neq (A_{>i})_{1j} = \gamma, \tag{4.1}$$

say. With $x_{i+j} = 1$ and $x_{i+k} = 0$ for the remaining positive $k \neq j$, we obtain

$$\mathbf{x}_{>i}^T A_{>i} \mathbf{x}_{>i} = 2\gamma x_{i+1} + (A_{>i})_{ii}$$

which takes both signs if $|x_{i+1}|$ is large. Thus A is indefinite, and we are led to the following algorithm for positive semidefiniteness (PSD).

PSD Test. As for the PD Test, but replace 1 by:

1. Test each a_{jj} and stop if < 0: q is not PSD. Stop if $i = n$: q is PSD. Stop if $a_{11} = 0 \neq a_{1j}$ for some j: q is not PSD. If each $a_{1j} = 0$, delete row and column one to obtain $A_{>1}$, and go to 3.

Corresponding determinant tests need caution, as the following shows:

Example 1. $A = \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix}$. Then $\alpha_1 = \alpha_2 = 0$ in (3.1), but q may be PSD or NSD, depending on the sign of δ.

A standard test (e.g., [**6**, p. 257]) is that q is PSD if *all* principal minors of A (obtained by deleting some subset of rows and corresponding columns) are non-negative. Although elegant, this implicitly amounts to permuting the x_i in $q(\mathbf{x})$, hence to permuting the rows and columns of A. Such tests are explicit in, for example, [**1**, §203] and [**2**], but are computationally prohibitive if n is large.

A complete definiteness classification is easily constructed for the PSD test, in contrast with most tests in the literature (though see [**6**]).

General Test. If the PSD test fails, either the pivot is negative, and then we may continue with Step 2, or else (4.1) holds. Thus either all the p_i may be obtained, and the definiteness status may be read from (1.2), or else at some stage (4.1) forces q to be indefinite.

5. Constrained pivot tests. The first objective is to determine a basis of solutions to $C\mathbf{x} = \mathbf{0}$, and this may be accomplished simply once C is converted to reduced row echelon form by elementary row operations. Explicitly, we may write

$$LC = [I \quad E]P \tag{5.1}$$

where L is invertible, $m \times m$, and lower triangular; I is $m \times m$; P is an $n \times n$ permutation matrix; and E is $m \times l$. Here we write $l = n - m$ and we recall that C is $m \times n$ with rank m. It follows that the general solution to (1.3) may be written in "parametric form" as

$$P\mathbf{x} = \begin{bmatrix} -E\mathbf{t} \\ \mathbf{t} \end{bmatrix}, \tag{5.2}$$

where $\mathbf{t} \in \mathbb{R}^l$ is a vector of parameters, and so

$$\mathbf{x} = N\mathbf{t} \quad \text{where } N = P\begin{bmatrix} -E \\ I \end{bmatrix}.$$

We may now substitute this into (1.1) to obtain an unconstrained quadratic $q(\mathbf{x}) = \mathbf{t}^T R\mathbf{t}$, in just l variables, where $R = N^T AN$. In order to calculate R efficiently, we may partition

$$P^T AP = \begin{bmatrix} F & G \\ G^T & H \end{bmatrix}, \tag{5.3}$$

where F is $m \times m$, G is $m \times l$ and H is $l \times l$. Then

$$R = E^T FE - E^T G - G^T E + H \tag{5.4}$$

and we have the following

Constrained Tests.

1. Calculate E and P from (5.1), and R from (5.3), (5.4).
2. Carry out the tests of Sections 3 and 4 with A replaced by R.

6. Comparison with border tests. Early tests were of eigenvalue type, and evolved into *bordered determinant* tests; see [**3**, pp. 114–116], [**2**] for further references, and [**2**, p. 297 footnote] for a version involving the top left $i \times i$ subdeterminants β_i of $B = \begin{bmatrix} 0 & C \\ C^T & A \end{bmatrix}$. Many recent authors have called this version the *bordered Hessian* test, since the accessory problems of Section 1 involve the Hessian A of the Lagrangian. It can be shown that

$$\rho_i = \beta_{i+2m}/\beta_{i+2m-1}, \qquad 1 \leqslant i \leqslant l, \tag{6.1}$$

where ρ_i are the pivots of R tested in Section 5. It follows that q is PD (resp. ND) if

$$0 < (-1)^m \beta_{i+2m} \left(\text{resp. } (-1)^{i+m} \beta_{i+2m}\right), \qquad 1 \leqslant i \leqslant l.$$

These sign conditions can confuse beginners, and even though Väliaho [**8**] removes the factor of $(-1)^m$ by using $\begin{bmatrix} 0 & C \\ -C^T & A \end{bmatrix}$ instead of B, the problem of motivation remains.

It might be noted that some proofs of (6.1) use (5.4), but the computational aspect seems to have been ignored, except by Luenberger [**5**, §10.6]. His method takes longer to compute than ours, requiring Gram–Schmidt orthonormalisation and eigenvalue calculations. On the other hand it gives more detailed information than definiteness, aimed at convergence rates for critical point search algorithms.

7. Zero pivots. If some pivots ρ_i vanish, then the determinant methods of Section 3 may still be applied to the analogue of R with the t_j permuted. Since $t_j = x_{m+j}$ by (5.2), this is tantamount to recalculating the β_{i+2m} of (6.1) but for a B obtained by permuting the x_{m+j}. Spring [**6**] has derived this form of the border test, improving previous tests (e.g., [**2**], [**8**]) which implicitly or explicitly require permutation of *all* the x_i. Even Spring's test is computationally prohibitive, however, if l is large. By contrast the tests of Section 5 are usually easier to calculate if some ρ_j vanish.

Let us illustrate this with the following.

Example 2 [**6**, Ex. 3]. Classify $q(\mathbf{x}) = -x_1^2 + x_2^2 + x_3^2 - x_4^2$ subject to $x_1 - x_3 = 0$. Spring calculates the top left 3×3, 4×4 and 5×5 minors of

$$B = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

followed by those obtained from permuting the final three rows and corresponding columns. Although these calculations are simple, they are numerous, and we emphasize that this is the most efficient of the border tests.

The test of Section 5 starts with $A = \operatorname{diag}(-1, 1, 1, -1)$ and $C = [1, 0, -1, 0]$ so

$$E = [0, -1, 0],\ P = I \quad \text{and} \quad P^T A P = A.$$

The partitioning (5.3) of $P^T A P$ gives

$$F = [-1],\ G = [0, 0, 0] \quad \text{and} \quad H = \operatorname{diag}(1, 1, -1)$$

so

$$E^T F E = \operatorname{diag}(0, -1, 0),\ G^T E = E^T G = 0$$

and finally (5.4) gives $R = \operatorname{diag}(1, 0, -1)$. The pivots ρ_i are therefore 1, 0 and -1 and so $q(\mathbf{x})$ is indefinite subject to the constraint. Only two nontrivial multiplications are required, both for $E^T F E$.

REFERENCES

1. C. Carathéodory, Calculus of Variations and Partial Differential Equations of the First Order, Chelsea, New York, 1982. (First published by Teubner, 1935.)
2. G. Debreu, Definite and semidefinite quadratic forms, *Econometrica*, 20 (1952) 295–300.
3. H. Hancock, Theory of Maxima and Minima, Dover, New York, 1960. (First published by Ginn, 1917.)
4. P. Lancaster, M. Tismenetsky, The Theory of Matrices: With Applications, Academic Press, New York, 1985.
5. D. G. Luenberger, Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass., 1984.
6. D. Spring, On the second derivative test for constrained local extrema, this MONTHLY, 92 (1985) 631–643.
7. G. Strang, Linear Algebra and its Applications, Academic Press, New York, 1980.
8. H. Väliaho, On the definity of quadratic forms to linear constraints, *J. Optimization Theory Appl.*, 38 (1982) 143–145.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, BRUCE C. BERNDT, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, MICHAEL A. FILASETA, IRA M. GESSEL, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, JOSEPH B. MILES, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before May 31, 1992 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3469. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

Suppose P is a point in the interior of triangle ABC and suppose AP, BP, CP meet the lines BC, CA, AB respectively at the points D, E, F. Prove that the centroids of the six triangles PBD, PDC, PCE, PEA, PAF, PFB lie on a conic if and only if P lies on at least one of the three medians of the triangle.

E 3470. *Proposed by B. M. M. de Weger, University of Twente, Enschede, The Netherlands.*

Let S_n be the nth Schröder number, defined as the number of polygonal paths in the Cartesian plane that start at $(0, 0)$, end at (n, n), contain no points above the line $y = x$, and are composed of steps taken from the set $\{(0, 1), (1, 0), (1, 1)\}$. E 3343 [1989, 734; 1991, 367] asserted that $S_n \equiv 0 \pmod 3$ when n is even. Determine S_n modulo 3 when n is odd.

E 3471. *Proposed by William Calbeck, Florida International University, Miami, FL, and Bruce Reznick, University of Illinois at Urbana-Champaign.*

Let P_k be the set of all integer valued polynomials of degree at most k, i.e., the set of all polynomials p of degree at most k such that $p(n) \in \mathbb{Z}$ for $n \in \mathbb{Z}$. (It is known that $p \in P_k$ if and only if

$$p(x) = a_0 + a_1\binom{x}{1} + a_2\binom{x}{2} + \cdots + a_k\binom{x}{k},$$

where $a_0, a_1, a_2, \ldots, a_k$ are integers.) Let $r(k)$ be the smallest power of 2 strictly greater than k.

(a) If $p \in P_k$, show that the sequence $\{(-1)^{p(n)}\}_{n=1}^{\infty}$ is periodic with period $r(k)$.

(b) Show that any given sequence of plus and minus ones with period 2^n occurs for some p in P_{2^n-1}

E 3472. *Proposed by Hunter Snevily, California Institute of Technology, Pasadena.*

Suppose h and k are relatively prime positive integers and $n = h + k$. Show that for each j there are $h^{-1}\binom{n-1}{k}$ k-element subsets of $\{1, 2, \ldots, n-1\}$ with sum congruent to j modulo h.

E 3473. *Proposed by Lawrence J. Wallen, University of Hawaii at Manoa, Honolula.*

Suppose $0 \leqslant \theta_1 < \theta_2 < \cdots < \theta_n < \pi$. Let A be the n by n matrix whose entry in the ith row and jth column is $\sin|\theta_i - \theta_j|$. Show that

$$\|A\| \leqslant \cot(\pi/2n)$$

and that the estimate is best possible. Here $\|A\|$ is defined as $\sup |AX|$, where the supremum is taken over all column vectors X in $\mathbb{R}^n$ with Euclidean norm 1.

E 3474. *Proposed by Andrew Lenard, Indiana University, Bloomington.*

Let S be the set of nonnegative real numbers whose expansion to the base 4 involves only the digits 0 and 1.

(a) Prove that if $x \in S$, $y \in S$, and $x \neq y$, then $(x + y)/2 \notin S$.

(b) Suppose T is a set of nonnegative real numbers properly containing S. Prove that there exist $x \in T$, $y \in T$, $x \neq y$ such that $(x + y)/2 \in T$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Path-Connected Homeomorphic Decompositions of the Disc

E 3220 [1987, 680]. *Proposed by Erwin Kronheimer, Birkbeck College, University of London.*

(a) Is it possible to decompose a closed 2-disc into two disjoint homeomorphic path-connected subsets?

(b) Into four such sets?

(The corresponding question for three sets seems difficult.)

Solution I by the proposer. Let $D = \{(x, y) | x^2 + y^2 \leqslant 1\}$ and let A be the union of the open upper half disc $\{(x, y) | x^2 + y^2 < 1 \text{ and } y > 0\}$ with part of its boundary, as described below.

In mostly counterclockwise order, the part consists of a point P on the open diameter $\{(x, 0) | -1 < x < 1\}$; a sequence of disjoint open intervals G_n for $n \geqslant 1$ on that diameter, converging monotonically to P from the right as $n \to \infty$; the open semicircle $G_0 = \{(x, y) | x^2 + y^2 = 1 \text{ and } y > 0\}$; a sequence of disjoint open intervals G_{-n} for $n \geqslant 1$ on the diameter, converging monotonically to a point P' from the left, where P' is to the left of P; a sequence of disjoint closed intervals

Slowly Converging Sets

E 3367 [1990, 150]. *Proposed by John Kieffer, University of Minnesota, Minneapolis.*

Let A_0 and A_1 be disjoint infinite sets of positive integers whose union is the set Z^+ of all positive integers. For each infinite subset S of Z^+ let $f(S)$ be the set obtained from S by removing the least element s of S and then adjoining all the elements of A_{1-i} less than s if s is in A_i. Let $f^2(S) = f(f(S))$, $f^3(S) = f(f^2(S))$, etc. Given any positive integer k, prove that for sufficiently large n either

$$k \in f^n(A_0) \cap f^n(A_1) \quad \text{or} \quad k \notin f^n(A_0) \cup f^n(A_1).$$

Solution by David Callan, University of Wisconsin, Madison, and Douglas B. West, University of Illinois, Urbana. For a given partition of Z^+ into two infinite sets A and B, we obtain the stronger result of computing the exact minimum value $g(k)$ such that $f^n(A)$ and $f^n(B)$ agree on k (both contain or both omit) for all $n \geqslant g(k)$. In terms of k alone, $g(k)$ is bounded by the Fibonacci number F_k, define by

$$F_0 = 1,\ F_1 = 2,\ F_k = F_{k-1} + F_{k-2} \quad \text{for } k \geqslant 2.$$

Let C_j be the jth maximal block of consecutive integers intersecting only one of $\{A, B\}$, written in increasing order. Using concatenation to write sets in increasing order, we have $Z^+ = C_1C_2C_3C_4 \cdots$, and we may assume $A = C_1C_3 \cdots$ and $B = C_2C_4 \cdots$. Let $c_j = |C_j|$, and let z_j be the largest element in C_j; we compute $g(k)$ from k and $c_1, \ldots, c_j$, where $k \in C_j$. Let $m(S)$ denote the smallest element of a set S, and let $S' = S - \{m(S)\}$. Let $D_j = C_{j'} \cdots C_{j-2}C_j$, where $j' = j - 2\lfloor j/2 \rfloor$ and $C_0 = \varnothing (z_0 = 0)$. We have $f(S) = D_jS'$ for any set S with $m(S) \in C_{j+1}$.

Define a sequence a_j by $a_{-1} = a_0 = 1$ and $a_j = a_{j-2} + c_ja_{j-1}$ for $j \geqslant 1$. For any set S with $m(S) \in C_{j+1}$, we claim $f^{a_j}(S) = S'$ and, further, each intermediate set $f^i(S)$ intersects $[m(C_{j+1}), \infty)$ in S', and the sequence of initial segments $\{f^i(S) \cap C_1C_2 \cdots C_j\}_{1 \leqslant i < a_j}$ depends only on j. For $j = 0$, this is clear. Suppose $j \geqslant 1$ and $m(S) \in C_{j+1}$. We have $f(S) = D_jS' = f(m(C_{j-1})C_jS')$ [$m(C_0)$ vanishing] and so by the inductive hypothesis, $f^{a_j}(S) = f^{c_ja_{j-1}+a_{j-2}}(m(C_{j-1})C_jS') = f^{c_ja_{j-1}}(C_jS') = S'$. The other statements also hold by induction on j since they are true for $i = 1$. It follows that the set $U = U(S, k) = \{n\colon 1 \leqslant n \leqslant a_j,\ k \in f^n(S)\}$ depends only on k for $k \leqslant z_j$ and is empty for $z_j < k \leqslant m(S)$.

Now let $E_j = C_jC_{j+2}C_{j+4} \cdots$. We claim that $\{A, B\}$ reaches $\{z_jE'_{j+1}, E_{j+2}\}$ by $a_j - 1$ simultaneous iterations, and we track the iterates to ascertain on which k they agree. For $j = 1$, $f^{c_1}(A) = E_3$ and $f^{c_1}(B) = z_1E'_2$ as desired. We have $(z_jE'_{j+1}, E_{j+2}) \to (E'_{j+1}, C_{j+1}E'_{j+2})$ after a_{j-1} iterations. The former pair disagrees on z_j and the latter pair on $z_j + 1$. After ia_j further steps, for $1 \leqslant i < c_{j+1}$, the resulting pair disagrees on $z_j + 1 + i$ and the final pair is $(E_{j+3}, z_{j+1}E'_{j+2})$. This is because each of these groups of a_j steps erases the current first term of each set, by the first result of the previous paragraph. (Note the reversal of order in the final pair.) Thus by the inductive hypothesis $\{A, B\}$ reaches $\{z_{j+1}E'_{j+2}, E_{j+3}\}$ in $(a_j - 1) + a_{j-1} + (c_{j+1} - 1)a_j = a_{j+1} - 1$ steps as required. Observe that the entire sequence $(z_jE'_{j+1}, E_{j+2}) \to (E_{j+3}, z_{j+1}E'_{j+2})$ agrees on all $k < z_j$: for the first a_{j-1} steps by inspection, thereafter by the results of the preceding paragraph. Further,

agreement holds (for the same reasons) on $z_j + i$ at all pairs following the noted disagreement, for $0 \leqslant i < c_j$. This determines g and yields $g(z_j) = a_j$, $g(z_j + i) = a_{j-1} + ia_j$ for $1 \leqslant i \leqslant c_{j+1}$.

Finally, we prove $g(k) \leqslant F_k$ for every partition A, B. From the formula $g(z_j + i) = a_{j-1} + ia_j$, we have $g(k) = g(k-1) + a_j$ if $z_j + 1 < k \leqslant z_{j+1}$, and $g(k) = g(k-1) + a_{j-1}$ if $k = z_j + 1$. Since $a_j = g(z_j)$ for all j and the sequence g is increasing, we have $g(k) \leqslant g(k-1) + g(k-2)$, with equality if and only if $k = z_j + 2$ or $k = z_j + 1 = z_{j-1} + 2$. In particular, $g(k) = F_k$ if and only if $c_j = 1$ for $1 \leqslant j \leqslant k - 2$.

The original problem was solved by R. J. Hendel, Central Michigan University Problem Group, Western Maryland College Problems Group, and the proposer. Two incomplete solutions were also received.

Reversing a Recurrence Relation

E 3374 [1990, 239]. *Proposed by N. J. Fine, Deerfield Beach, FL.*

Suppose J is a function on the non-negative integers such that $J(0) = J(1) = 2$ and

$$J(n+1) = J(n) + \lfloor J(n-1)/2 \rfloor \quad \text{for } n \geqslant 1.$$

Prove that

$$\left\lfloor (\sqrt{3} - 1)(J(n) + 1) \right\rfloor = J(n-1) \quad \text{for } n \geqslant 1.$$

Solution by David Callan, University of Wisconsin, Madison. The result follows by induction on n, using the following lemma.

LEMMA. *If b is a positive integer, $a = \lfloor (b+1)(\sqrt{3} - 1) \rfloor$, and $c = b + \lfloor a/2 \rfloor$, then $\lfloor (c+1)(\sqrt{3} - 1) \rfloor = b$.*

Proof. By hypothesis, $a < (b+1)(\sqrt{3} - 1) < a + 1$. Since a is an integer, $b + (a-1)/2 \leqslant c \leqslant b + a/2$. The inequalities $c \leqslant b + a/2$ and $a < (b+1) \times (\sqrt{3} - 1)$ imply $c + 1 < (b/2)(\sqrt{3} + 1) + (\sqrt{3} + 1)/2$, which upon multiplication by $\sqrt{3} - 1$ transforms to $(c+1)(\sqrt{3} - 1) < b + 1$.

The inequalities $(b+1)(\sqrt{3} - 1) < a + 1$ and $b + (a+1)/2 \leqslant c + 1$ imply $(b/2)(\sqrt{3} + 1) + (\sqrt{3} - 1)/2 < c + 1$, which by the same multiplication as before transforms to $b + (\sqrt{3} - 1)^2/2 < (c+1)(\sqrt{3} - 1)$. Hence $b < (c+1)(\sqrt{3} - 1)$, which completes the proof.

Solved also by J. Anglesio (France), R. J. Chapman (Great Britain), D. Doster, J. Guillera (Spain), C. Hill, A. A. Jagers (The Netherlands), P. Karson (student), O. P. Lossers (The Netherlands), L. E. Mattics, T. S. Norfolk, M. Riazi-Kermani, H. D. Ruderman, J. H. Steelman, R. Stong, J. S. Summer, M. Vowe (Switzerland), J. T. Ward, H. C. Wei, and the Central Michigan University Problem Group. One partial solution was submitted.

A Complex Sequence That Is Nonzero Infinitely Often

E 3384 [1990, 427]. *Proposed by Jacob Sturm, Rutgers University, Newark, NJ.*

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of distinct complex numbers such that $\sum |a_n|$ converges. Let $\{c_n\}_{n=1}^{\infty}$ be a bounded sequence of non-zero complex numbers. Prove

agreement holds (for the same reasons) on $z_j + i$ at all pairs following the noted disagreement, for $0 \leqslant i < c_j$. This determines g and yields $g(z_j) = a_j$, $g(z_j + i) = a_{j-1} + ia_j$ for $1 \leqslant i \leqslant c_{j+1}$.

Finally, we prove $g(k) \leqslant F_k$ for every partition A, B. From the formula $g(z_j + i) = a_{j-1} + ia_j$, we have $g(k) = g(k-1) + a_j$ if $z_j + 1 < k \leqslant z_{j+1}$, and $g(k) = g(k-1) + a_{j-1}$ if $k = z_j + 1$. Since $a_j = g(z_j)$ for all j and the sequence g is increasing, we have $g(k) \leqslant g(k-1) + g(k-2)$, with equality if and only if $k = z_j + 2$ or $k = z_j + 1 = z_{j-1} + 2$. In particular, $g(k) = F_k$ if and only if $c_j = 1$ for $1 \leqslant j \leqslant k - 2$.

The original problem was solved by R. J. Hendel, Central Michigan University Problem Group, Western Maryland College Problems Group, and the proposer. Two incomplete solutions were also received.

Reversing a Recurrence Relation

E 3374 [1990, 239]. *Proposed by N. J. Fine, Deerfield Beach, FL.*

Suppose J is a function on the non-negative integers such that $J(0) = J(1) = 2$ and

$$J(n+1) = J(n) + \lfloor J(n-1)/2 \rfloor \quad \text{for } n \geqslant 1.$$

Prove that

$$\left\lfloor (\sqrt{3} - 1)(J(n) + 1) \right\rfloor = J(n-1) \quad \text{for } n \geqslant 1.$$

Solution by David Callan, University of Wisconsin, Madison. The result follows by induction on n, using the following lemma.

LEMMA. *If b is a positive integer, $a = \lfloor (b+1)(\sqrt{3} - 1) \rfloor$, and $c = b + \lfloor a/2 \rfloor$, then $\lfloor (c+1)(\sqrt{3} - 1) \rfloor = b$.*

Proof. By hypothesis, $a < (b+1)(\sqrt{3} - 1) < a + 1$. Since a is an integer, $b + (a-1)/2 \leqslant c \leqslant b + a/2$. The inequalities $c \leqslant b + a/2$ and $a < (b+1) \times (\sqrt{3} - 1)$ imply $c + 1 < (b/2)(\sqrt{3} + 1) + (\sqrt{3} + 1)/2$, which upon multiplication by $\sqrt{3} - 1$ transforms to $(c+1)(\sqrt{3} - 1) < b + 1$.

The inequalities $(b+1)(\sqrt{3} - 1) < a + 1$ and $b + (a+1)/2 \leqslant c + 1$ imply $(b/2)(\sqrt{3} + 1) + (\sqrt{3} - 1)/2 < c + 1$, which by the same multiplication as before transforms to $b + (\sqrt{3} - 1)^2/2 < (c+1)(\sqrt{3} - 1)$. Hence $b < (c+1)(\sqrt{3} - 1)$, which completes the proof.

Solved also by J. Anglesio (France), R. J. Chapman (Great Britain), D. Doster, J. Guillera (Spain), C. Hill, A. A. Jagers (The Netherlands), P. Karson (student), O. P. Lossers (The Netherlands), L. E. Mattics, T. S. Norfolk, M. Riazi-Kermani, H. D. Ruderman, J. H. Steelman, R. Stong, J. S. Summer, M. Vowe (Switzerland), J. T. Ward, H. C. Wei, and the Central Michigan University Problem Group. One partial solution was submitted.

A Complex Sequence That Is Nonzero Infinitely Often

E 3384 [1990, 427]. *Proposed by Jacob Sturm, Rutgers University, Newark, NJ.*

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of distinct complex numbers such that $\sum |a_n|$ converges. Let $\{c_n\}_{n=1}^{\infty}$ be a bounded sequence of non-zero complex numbers. Prove

(b) Find the largest constant c such that

$$\left(\frac{\sin x}{x}\right)^2 + \frac{\tan x}{x} > 2 + cx^3 \tan x$$

for $0 < x < \pi/2$.

Editorial remark. The editors overlooked the fact that the following solution of (b) is much shorter than the published solution by Jean Anglesio. Other solutions of (b) were received from Richard A. Groeneveld, Kee-Wai Lau (Hong Kong), J. Ernest Wilkins, Jr., and the Lamar University Problem Solving Group.

Solution II of (b) by David Callan, University of Wisconsin, Madison. We must determine the infimum on the interval $(0, \pi/2)$ of the function ϕ defined on the interval $(0, \pi)$ by

$$(*) \qquad \phi(x) = \left\{(x^{-1}\sin x)^2 + x^{-1}\tan x - 2\right\}/\{x^3 \tan x\}$$

$$= 2^{-1}x^{-5}\sin 2x + x^{-4} - 2x^{-3}\cot x.$$

We claim that ϕ is decreasing on $(0, \pi/2)$, so that

$$\inf_{0<x<\pi/2} \phi(x) = \phi(\pi/2) = 16/\pi^4.$$

Thus $16/\pi^4$ is the largest value of c satisfying (b).

On the interval $(0, \pi)$ we have the two Maclaurin expansions

$$\sin 2x = 2x - 4x^3/3 + \sum_{n=0}^{\infty} (-1)^n (2x)^{2n+5}/(2n+5)!,$$

$$x\cot x = 1 - x^2/3 - \sum_{n=0}^{\infty} 2^{2n+4}|B_{2n+4}|x^{2n+4}/(2n+4)!;$$

the latter is easily deduced from the fact that the Bernoulli numbers B_r can be defined by the power series expansion $t/(e^t - 1) = \sum_{r=0}^{\infty} B_r t^r/r!$ for $|t| < 2\pi$. Hence for $0 < x < \pi$ we have

$$\phi(x) = \sum_{n=0}^{\infty} \frac{2^{2n+5}}{(2n+4)!}\left\{|B_{2n+4}| + \frac{(-1)^n}{2(2n+5)}\right\}x^{2n}.$$

Thus, if $\psi(x) = \phi(\sqrt{x})$ for $0 < x < \pi^2$, we have

$$\psi(x) = \frac{8}{45} - \frac{8x}{945} + \frac{16x^2}{14175} + \sum_{n=3}^{\infty} \frac{2^{2n+5}}{(2n+4)!}\left\{|B_{2n+4}| + \frac{(-1)^n}{4n+10}\right\}x^n.$$

Since

$$2(2n+5)|B_{2n+4}| = 4(2n+5)!(2\pi)^{-2n-4}\sum_{k=1}^{\infty} k^{-2n-4}$$

$$> 4(2n+5)!(2\pi)^{-2n-4} > 1$$

for $n \geqslant 3$, we have $|B_{2n+4}| > 1/(4n+10)$ for $n \geqslant 3$. Thus $\psi''(x) > 0$ on the interval $(0, \pi^2)$ and so ψ' is increasing there. On the other hand, from $(*)$ we

readily obtain

$$\psi'\left((\pi/2)^2\right) = \pi^{-1}\phi'(\pi/2) = (2/\pi)^4(1 - 10/\pi^2) < 0.$$

It follows that $\psi'(x)$ is negative throughout the interval $(0,(\pi/2)^2)$ and so ψ is decreasing there. Hence ϕ is decreasing on the interval $(0, \pi/2)$ and our claimed result is established.

Rationally Independent Subsets of the Reals

E 3319 [1989, 356; 1990, 854]. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest, and János Surányi, Eötvös Loránd University, Budapest.*

Suppose S is a set of $2n + 1$ irrational real numbers. Prove that S has a subset T of cardinality $n + 1$ such that no nonempty subset of T has a rational sum.

Editorial remark. Professor Günter Rote of the Technical University of Graz, Austria, has pointed out to us that the second solution published in the November, 1990 issue is flawed, in that the hyperplane mentioned in the solution should contain the set $\mathbb{Q}$ of rational numbers rather than merely the origin 0. A corrected version of this solution is as follows.

Solution II by an anonymous contributor. Let V be the finite-dimensional vector space over $\mathbb{Q}$ generated by 1 and the elements of S. Since S is a finite set of real numbers none of which is contained in $\mathbb{Q}$, there exists a hyperplane H through the one-dimensional subspace of V generated by 1 such that H contains no elements of S. Clearly $n + 1$ of the elements of S fall on the same side of H. These $n + 1$ elements of S form a set T having the property stated in the problem. In fact no nontrivial linear combination of the elements of T with nonnegative rational coefficients is in H, so that in particular no non-empty subset of T has a rational sum.

Sum-Free Sets Modulo n

E 3346 [1989, 735; 1991, 368]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

Call a subset T of $\mathbb{Z}$ mod n *sum-free* if the sum of two distinct elements of T is not in T. Let $s(n)$ be the maximum cardinality of a sum-free subset of $\mathbb{Z}$ mod n.

(a) Prove that $s(n) = n/2$ if n is even.

(b) If n is odd, prove that $s[n] \geqslant \lfloor n/3 \rfloor + 1$.

(c)* If n is odd, is $s(n) = \lfloor n/3 \rfloor + 1$?

Editorial Comment. For odd n several solvers (correctly) conjectured that "$s(n) = \lfloor n/3 \rfloor + 1$ if n has no prime factors congruent to 5 modulo 6, but $s(n) = n(p + 1)/(3p)$ if p is the smallest prime factor of n congruent to 5 modulo 6."

Fred Galvin has remarked to us that the preceding conjecture follows easily from the following known result: If $s^*(n)$ is the maximum cardinality of a subset T of $\mathbb{Z}$ mod n which does not contain the sum of any two elements of T, *distinct or not*, then

$$(*) \qquad s^*(n) = \max_{d|n} \left\{ \left\lfloor \frac{d+1}{3} \right\rfloor \frac{n}{d} \right\}.$$

readily obtain

$$\psi'\left((\pi/2)^2\right) = \pi^{-1}\phi'(\pi/2) = (2/\pi)^4(1 - 10/\pi^2) < 0.$$

It follows that $\psi'(x)$ is negative throughout the interval $(0,(\pi/2)^2)$ and so ψ is decreasing there. Hence ϕ is decreasing on the interval $(0, \pi/2)$ and our claimed result is established.

Rationally Independent Subsets of the Reals

E 3319 [1989, 356; 1990, 854]. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest, and János Surányi, Eötvös Loránd University, Budapest.*

Suppose S is a set of $2n + 1$ irrational real numbers. Prove that S has a subset T of cardinality $n + 1$ such that no nonempty subset of T has a rational sum.

Editorial remark. Professor Günter Rote of the Technical University of Graz, Austria, has pointed out to us that the second solution published in the November, 1990 issue is flawed, in that the hyperplane mentioned in the solution should contain the set $\mathbb{Q}$ of rational numbers rather than merely the origin 0. A corrected version of this solution is as follows.

Solution II by an anonymous contributor. Let V be the finite-dimensional vector space over $\mathbb{Q}$ generated by 1 and the elements of S. Since S is a finite set of real numbers none of which is contained in $\mathbb{Q}$, there exists a hyperplane H through the one-dimensional subspace of V generated by 1 such that H contains no elements of S. Clearly $n + 1$ of the elements of S fall on the same side of H. These $n + 1$ elements of S form a set T having the property stated in the problem. In fact no nontrivial linear combination of the elements of T with nonnegative rational coefficients is in H, so that in particular no non-empty subset of T has a rational sum.

Sum-Free Sets Modulo n

E 3346 [1989, 735; 1991, 368]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

Call a subset T of $\mathbb{Z}$ mod n *sum-free* if the sum of two distinct elements of T is not in T. Let $s(n)$ be the maximum cardinality of a sum-free subset of $\mathbb{Z}$ mod n.

(a) Prove that $s(n) = n/2$ if n is even.

(b) If n is odd, prove that $s[n] \geqslant \lfloor n/3 \rfloor + 1$.

(c)* If n is odd, is $s(n) = \lfloor n/3 \rfloor + 1$?

Editorial Comment. For odd n several solvers (correctly) conjectured that "$s(n) = \lfloor n/3 \rfloor + 1$ if n has no prime factors congruent to 5 modulo 6, but $s(n) = n(p + 1)/(3p)$ if p is the smallest prime factor of n congruent to 5 modulo 6."

Fred Galvin has remarked to us that the preceding conjecture follows easily from the following known result: If $s^*(n)$ is the maximum cardinality of a subset T of $\mathbb{Z}$ mod n which does not contain the sum of any two elements of T, *distinct or not*, then

$$(*) \qquad s^*(n) = \max_{d|n} \left\{ \left\lfloor \frac{d+1}{3} \right\rfloor \frac{n}{d} \right\}.$$

Proofs of (*) may be found in [1] and [2]. To deduce the assertion in quotation marks from (*), it suffices to prove the following lemma given by Galvin.

LEMMA. *If T is a sum-free set of $\mathbb{Z} \bmod n$ in the sense of the problem and if $2a \in T$ for some $a \in T$, then $|T| \leqslant \lfloor n/3 \rfloor + 1$.*

Proof. If $0 \in T$, then $T = \{0\}$. Thus we may assume $a \neq 0$. Let $T = \{a, 2a, x_1, \ldots, x_{t-2}\}$, where $t = |T|$. Then the $2(t-2)+1$ elements $3a, a + x_1, \ldots, a + x_{t-2}, 2a + x_1, \ldots, 2a + x_{t-2}$ are all different and none of them belongs to T. Hence $2(t-2)+1 \leqslant n - t$, i.e., $t \leqslant n/3 + 1$.

It follows from the lemma that $s(n) = \max\{s^*(n), \lfloor n/3 \rfloor + 1\}$. This gives the following evaluations. If n is even, $s(n) = s^*(n) = n/2$. If n is odd and has no prime factor congruent to 5 modulo 6, then $s^*(n) = \lfloor n/3 \rfloor$ and $s(n) = \lfloor n/3 \rfloor + 1$. If n is odd and has a prime factor congruent to 5 modulo 6, then $s(n) = s^*(n) = (p+1)n/(3p)$, where p is the smallest such prime factor.

REFERENCES

1. Palahenedi Hewage Diananda and Hian Poh Yap, Maximal sum-free sets of elements of finite groups, *Proc. Japan Acad.*, 45 (1969) 1–5.
2. Anne Penfold Street, Sum-free sets, *Springer Lecture Notes in Mathematics*, 292 (1972) 123–271.

ADVANCED PROBLEMS

6673. *Proposed by Paresh J. Malde and Allen J. Schwenk, Western Michigan University, Kalamazoo.*

Let $c_n = e_n - d_n$, where e_n is the number of connected labeled n-vertex graphs with an even number of edges and d_n is the number of connected labeled n-vertex graphs with an odd number of edges. For example, $c_2 = 0 - 1 = -1$ and $c_3 = 3 - 1 = 2$. Find a general formula of c_n.

6674. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

If n is an integer greater than 1, let $P(n)$ denote the largest prime factor of n. Prove that $n|P(n)!$ for almost all n, i.e., prove that if

$$S(x) = \{n \leqslant x : n \nmid P(n)!\},$$

then

$$\lim_{x \to \infty} |S(x)|/x = 0.$$

6675. *Proposed by Alain Tissier, Montfermeil, France.*

Define

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^{1/n}}{n}$$

for $x > 0$. Determine the asymptotic behavior of $f(x)$ as $x \to 0$.

Proofs of (*) may be found in [1] and [2]. To deduce the assertion in quotation marks from (*), it suffices to prove the following lemma given by Galvin.

LEMMA. *If T is a sum-free set of $\mathbb{Z} \bmod n$ in the sense of the problem and if $2a \in T$ for some $a \in T$, then $|T| \leqslant \lfloor n/3 \rfloor + 1$.*

Proof. If $0 \in T$, then $T = \{0\}$. Thus we may assume $a \neq 0$. Let $T = \{a, 2a, x_1, \ldots, x_{t-2}\}$, where $t = |T|$. Then the $2(t-2)+1$ elements $3a, a + x_1, \ldots, a + x_{t-2}, 2a + x_1, \ldots, 2a + x_{t-2}$ are all different and none of them belongs to T. Hence $2(t-2)+1 \leqslant n - t$, i.e., $t \leqslant n/3 + 1$.

It follows from the lemma that $s(n) = \max\{s^*(n), \lfloor n/3 \rfloor + 1\}$. This gives the following evaluations. If n is even, $s(n) = s^*(n) = n/2$. If n is odd and has no prime factor congruent to 5 modulo 6, then $s^*(n) = \lfloor n/3 \rfloor$ and $s(n) = \lfloor n/3 \rfloor + 1$. If n is odd and has a prime factor congruent to 5 modulo 6, then $s(n) = s^*(n) = (p+1)n/(3p)$, where p is the smallest such prime factor.

REFERENCES

1. Palahenedi Hewage Diananda and Hian Poh Yap, Maximal sum-free sets of elements of finite groups, *Proc. Japan Acad.*, 45 (1969) 1–5.
2. Anne Penfold Street, Sum-free sets, *Springer Lecture Notes in Mathematics*, 292 (1972) 123–271.

ADVANCED PROBLEMS

6673. *Proposed by Paresh J. Malde and Allen J. Schwenk, Western Michigan University, Kalamazoo.*

Let $c_n = e_n - d_n$, where e_n is the number of connected labeled n-vertex graphs with an even number of edges and d_n is the number of connected labeled n-vertex graphs with an odd number of edges. For example, $c_2 = 0 - 1 = -1$ and $c_3 = 3 - 1 = 2$. Find a general formula of c_n.

6674. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

If n is an integer greater than 1, let $P(n)$ denote the largest prime factor of n. Prove that $n|P(n)!$ for almost all n, i.e., prove that if

$$S(x) = \{n \leqslant x : n \nmid P(n)!\},$$

then

$$\lim_{x \to \infty} |S(x)|/x = 0.$$

6675. *Proposed by Alain Tissier, Montfermeil, France.*

Define

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^{1/n}}{n}$$

for $x > 0$. Determine the asymptotic behavior of $f(x)$ as $x \to 0$.

SOLUTIONS OF ADVANCED PROBLEMS

A Set of Points Related to Nonnegative Cosine Polynomials

6608 [1989, 652]. *Proposed by Oliver D. Anderson, University of Western Ontario, London, Canada.*

Let S_n be the set of all points $(x_1, \ldots, x_n)$ in R^n such that

$$x_k = \sum_{j=0}^{n-k} t_j t_{j+k} \Big/ \sum_{j=0}^{n} t_j^2 \qquad (k = 1, 2, \ldots, n),$$

where $t_0, t_1, \ldots, t_n$ are arbitrary real numbers not all zero. It is easy to see that S_n is bounded and contains the origin as an interior point.

(a) When $n = 2$ prove that S_n is the convex hull of the set consisting of the point $(0, -\frac{1}{2})$ and the points of the ellipse $2x_1^2 + (4x_2 - 1)^2 = 1$.

(b)* Is S_n convex for all n?

(c)* Is S_n closed for all n?

Editorial remark. The version of the problem originally published specified that $t_0 = 1$. The symmetrical version given above is more convenient to work with. The two versions are equivalent. For suppose $P(t_0, t_1, \ldots, t_n)$ is the point in $\mathbb{R}^n$ whose kth coordinate is

$$\sum_{j=0}^{n-k} t_j t_{j+k} \Big/ \sum_{j=0}^{n} t_j^2 \qquad (k = 1, 2, \ldots, n),$$

where $t_0, t_1, \ldots, t_n$ are real numbers not all zero; then

$$P_n(t_0, t_1, \ldots, t_n) = P_n(1, t_1/t_0, t_2/t_0, \ldots, t_n/t_0) \quad \text{if } t_0 \neq 0$$

and

$$P_n(t_0, t_1, \ldots, t_n) = P_n(1, t_{r+1}/t_r, t_{r+2}/t_r, \ldots, t_n/t_r, 0, \ldots, 0) \quad \text{if } t_1 = \cdots = t_{r-1} = 0, t_r \neq 0.$$

Solution to (b) and (c) by Zoltán Sasvári, Technische Universität, Dresden, Germany. The answer to both questions is yes. A lemma of L. Fejér and F. Riesz ([**1**], [**2**], [**3**]) asserts that a cosine polynomial $\lambda_0 + \lambda_1 \cos t + \cdots + \lambda_n \cos nt$ with real coefficients is nonnegative for all real t if and only if it can be expressed in the form

$$\lambda_0 + \lambda_1 \cos t + \cdots + \lambda_n \cos nt = |\mu_0 + \mu_1 e^{it} + \cdots + \mu_n e^{int}|^2,$$

where $\mu_0, \mu_1, \ldots, \mu_n$ are real numbers. In other words the set of nonnegative cosine polynomials $\lambda_0 + \lambda_1 \cos t + \cdots + \lambda_n \cos nt$ can be parametrized as follows:

$$\lambda_0 = \sum_{j=0}^{n} \mu_j^2, \qquad \lambda_r = 2 \sum_{j=0}^{n-r} \mu_j \mu_{j+r} \qquad (r = 1, 2, \ldots, n),$$

where $(\mu_0, \mu_1, \ldots, \mu_n) \in \mathbb{R}^{n+1}$.

The definition of S_n, the parametrization just given, and an obvious normalization show that $(x_1, x_2, \ldots, x_n) \in S_n$ if and only if $1 + 2\sum_{k=1}^{n} x_k \cos kt$ is nonnega-

tive for all real t, i.e., if and only if

$$1 + 2\,\mathrm{Re} \sum_{k=1}^{n} x_k z^k \geqslant 0 \qquad (*)$$

for all complex z of absolute value 1. The set of points $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying $(*)$ for all z of modulus 1 is obviously closed and convex, so that assertions (b) and (c) follow.

REFERENCES

1. L. Fejér, Über trigonometrische Polynome, *J. Reine Angew. Math.*, 146 (1916) 53–82, page 64.
2. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Springer-Verlag, 1925, Volume 2, Part VI, Problems 39–41.
3. F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Ungar Publishing, 1955, pp. 117–118.

Solution of (a) by the editors. Since the lines through $(-1/2, 0)$ with slopes ± 1 are tangent to the ellipse $2x_1^2 + (4x_2 - 1)^2 = 1$ at the points $(\pm 2/3, 1/6)$, it suffices to show that $(x_1, x_2) \in S_2$ if and only if $-1/2 \leqslant x_2 \leqslant 1/2$ and

$$|x_1| \leqslant x_2 + 1/2 \quad \text{if } -1/2 \leqslant x_2 \leqslant 1/6,$$
$$|x_1| \leqslant \left(4x_2 - 8x_2^2\right)^{1/2} \quad \text{if } 1/6 \leqslant x_2 \leqslant 1/2.$$

From the above solution of (b) and (c) we note that $(x_1, x_2) \in S_2$ if and only if

$$1 + 2x_1 \cos t + 2x_2 \cos 2t \geqslant 0 \qquad (*)$$

for all t in $[0, \pi)$. Putting $v = \tan^2(t/2)$, we see that $(*)$ holds for all t in $[0, \pi)$ if and only if

$$1 + 2x_1(1 - v)/(1 + v) + 2x_2\left\{2(1 - v)^2/(1 + v)^2 - 1\right\} \geqslant 0$$

for all nonnegative real v. Thus $(x_1, x_2) \in S_2$ if and only if

$$1 + 2x_1 + 2x_2 + 2(1 - 6x_2)v + (1 - 2x_1 + 2x_2)v^2 \geqslant 0$$

for all $v \geqslant 0$. Now if a, b, c are given real numbers, $av^2 + bv + c \geqslant 0$ for all nonnegative v if and only if $a \geqslant 0$, $c \geqslant 0$, and either $b \geqslant 0$ or $ac - b^2/4 \geqslant 0$. Thus $(x_1, x_2) \in S_2$ if and only if $1 + 2x_1 + 2x_2 \geqslant 0$, $1 - 2x_1 + 2x_2 \geqslant 0$, and either $1 - 6x_2 \geqslant 0$ or $(1 - 6x_2)^2 - (1 + 2x_2)^2 + 4x_1^2 \leqslant 0$. Thus $(x_1, x_2) \in S_2$ if and only if $|x_1| \leqslant x_2 + 1/2$ and either $x_2 \leqslant 1/6$ or $x_1^2 + 8x_2^2 - 4x_2 \leqslant 0$. The assertion of (a) follows.

Editorial comment. Anderson informs us that (b) and (c) also follow from a result well known to time-series analysts, namely that the sum of two independent moving average processes is also a moving average.

Parts (b) and (c) were solved also by V. M. Joshi and the proposer. Direct solutions of (a) (i.e., solutions not using the Fejér–Riesz characterization of nonnegative cosine polynomials) were obtained by Everett Dade and Bruno Haible.

Zeros of Sections of the Zeta Function

6627 [1990, 252]. *Proposed by G. Boese, Munich, Germany*

If n is a positive integer, put

$$\zeta_n(s) = \sum_{j=1}^{n} j^{-s}.$$

It is known that there are infinitely many n such that ζ_n has zeros in $\{s\colon \operatorname{Re} s \geqslant 1\}$. However, prove that

(i) if $n = 1, 2, 3, 4, 5$, then $\operatorname{Re} \zeta_n(s) > 0$ for $\operatorname{Re} s \geqslant 1$;

(ii) if $n = 6, 7, 8, 9, 10$, then $|\zeta_n(s)| > 0$ for $\operatorname{Re} s \geqslant 1$.

Solution of (i) *by the editors*. For $n = 1, 2, 3$ we have $\operatorname{Re} \zeta_n(s) \geqslant 1 - 1/2 - 1/3 = 1/6$ for $\operatorname{Re} s \geqslant 1$.

For the cases $n = 4, 5$ it is helpful to consider the function f defined by $f(s) = 1 + 2^{-s} + 4^{-s}$. If $\operatorname{Re} s \geqslant 1$, we have $2^{-s} = x + iy$, where x and y are real and $x^2 + y^2 \leqslant 1/4$. Hence for $\operatorname{Re} s \geqslant 1$ we have

$$\begin{aligned} \operatorname{Re} f(s) &= 1 + x + x^2 - y^2 = 7/8 + 2(x + 1/4)^2 - x^2 - y^2 \\ &\geqslant 7/8 - 1/4 = 5/8, \end{aligned}$$

with equality if and only if $x = -1/4$ and $x^2 + y^2 = 1/4$, i.e., if and only if $s = 1 + 2\pi in/\log 8$, where n is an integer not divisible by 3. Thus if $n = 4$ or 5, we have

$$\operatorname{Re} \zeta_n(s) \geqslant 5/8 - 1/3 - 1/5 = 11/120$$

for $\operatorname{Re} s \geqslant 1$.

Solution of (ii) *by the proposer*. The cases $n = 6, 8, 9, 10$ can be handled very simply. Since

$$(1 - 2^{-s})\zeta_6(s) = 1 + 3^{-s} + 5^{-s} - 8^{-s} - 10^{-s} - 12^{-s},$$

we have

$$|(1 - 2^{-s})\zeta_6(s)| \geqslant 1 - 3^{-1} - 5^{-1} - 8^{-1} - 10^{-1} - 12^{-1} = 19/120$$

for $\operatorname{Re} s \geqslant 1$. Since

$$(1 - 2^{-s})\zeta_8(s) = 1 + 3^{-s} + 5^{-s} + 7^{-s} - 10^{-s} - 12^{-s} - 14^{-s} - 16^{-s}$$

we have

$$\begin{aligned} |(1 - 2^{-s})\zeta_8(s)| \geqslant\ & 1 - 3^{-1} - 5^{-1} - 7^{-1} \\ & - 10^{-1} - 12^{-1} - 14^{-1} - 16^{-1} = 11/1680 \end{aligned}$$

for $\operatorname{Re} s \geqslant 1$. Similarly, it is easy to check that the sum of the absolute values of the 15 terms of the Dirichlet series for

$$(1 - 2^{-s})(1 - 3^{-s})\zeta_9(s) - 1$$

is at most $1249/1260$ for $\operatorname{Re} s \geqslant 1$, so that

$$|(1 - 2^{-s})(1 - 3^{-s})\zeta_9(s)| \geqslant 1 - 1249/1260 = 11/1260$$

for $\operatorname{Re} s \geqslant 1$. Finally the sum of the absolute values of the 15 terms of the

Dirichlet series for

$$(1 - 2^{-s})(1 - 3^{-s})\zeta_{10}(s) - 1$$

is at most $233/252$ for $\operatorname{Re} s \geqslant 1$, so that

$$|(1 - 2^{-s})(1 - 3^{-s})\zeta_{10}(s)| \geqslant 1 - 233/252 = 19/252$$

for $\operatorname{Re} s \geqslant 1$.

The case $n = 7$ is less straightforward. Let

$$Z(u, v, w, t) = 1 + u + u^2 + v(1 + u) + w + t.$$

Since $Z(2^{-s}, 3^{-s}, 5^{-s}, 7^{-s}) = \zeta_7(s)$, it suffices to prove that $|Z(u, v, w, t)| > 0$ in the 4-disc

$$\{(u, v, w, t): |u| \leqslant 1/2, |v| \leqslant 1/3, |w| \leqslant 1/5, |t| \leqslant 1/7\}.$$

For fixed values of v, w, t with $|v| \leqslant 1/3$, $|w| \leqslant 1/5$, $|t| \leqslant 1/7$ we shall use Rouché's theorem to compare the number of zeros of $Z(u, v, w, t)$ in $\{u: |u| \leqslant 1/2\}$ with the number of zeros of $Z(u, 0, 0, 0) = 1 + u + u^2$ there. To apply Rouché's theorem it suffices to know that for $|u| = 1/2$ we have

$$|v(1 + u) + w + t| \leqslant |1 + u|/3 + 1/5 + 1/7 < |1 + u + u^2|.$$

If we put $x = 2 \operatorname{Re} u = u + \bar{u}$, where $u\bar{u} = 1/4$, then

$$|1 + u|^2 = 1 + u + \bar{u} + u\bar{u} = x + 5/4$$

and

$$\begin{aligned} |1 + u + u^2|^2 &= 1 + (u + \bar{u}) + (u + \bar{u})^2 - u\bar{u} + (u + \bar{u})u\bar{u} + (u\bar{u})^2 \\ &= (x + 5/8)^2 + 27/64. \end{aligned}$$

Thus it suffices to show that if

$$F(x) = \{(x + 5/8)^2 + 27/64\}^{1/2} - \{x + 5/4\}^{1/2}/3 - 12/35,$$

then $F(x) > 0$ for $x \in [-1, 1]$.

The positivity of F on $[-1, 1]$ can be verified in many different ways, including the following: (a) The minimum of F on $[-1, 1]$ can be determined by differentiating and then solving the resulting cubic equation. (b) Since it is easy to show that $|F'(x)| < 1$ for $x \in [-1, 1]$, it suffices to calculate that $F(x) > 0.02$ whenever x is an integer multiple of 0.04 in $[-1, 1]$. (c) By two squarings the inequality $F(x) > 0$ for $x \in [-1, 1]$ can easily be shown to be equivalent to the positivity of a certain quartic polynomial on $[-1, 1]$.

Once the positivity of F on $[0, 1]$ has been demonstrated, Rouché's theorem shows that for fixed v, w, t the function $Z(u, v, w, t)$ has the same number of zeros in $\{u: |u| \leqslant 1/2\}$ as the function $Z(u, 0, 0, 0)$, namely none. Thus $|Z(u, v, w, t)|$ has a positive lower bound for $|u| \leqslant 1/2$, $|v| \leqslant 1/3$, $|w| \leqslant 1/5$, $|t| \leqslant 1/7$ and so $|\zeta_7(s)|$ has a positive lower bound for $\operatorname{Re} s \geqslant 1$.

Editorial comment. Most of the results of this problem were obtained in a more computational manner by Robert Spira, "Zeros of Sections of the Zeta Function II," *Math. Comp.*, 22 (1960) 163–173. Spira also showed that ζ_{19} does have zeros with real part greater than 1.

No other solutions were received.

Schmidt's Query Becomes Lindsey's Theorem

6636 [1990, 535]. *Proposed by F. Schmidt, Bryn Mawr College, Bryn Mawr, PA.*

If G is a finite group, let $f(G)$ be the average order of an element of G, i.e.,

$$f(G) = |G|^{-1} \sum_{g \in G} \operatorname{ord}(g),$$

where $\operatorname{ord}(g)$ is the order of the group element g and $|G|$ is the number of elements in G.

(i) Show that if G is a nilpotent group of order n, then $f(G) \leqslant f(Z_n)$, where Z_n is the cyclic group of order n.

(ii)* Is it true that if G is any group of order n, then $f(G) \leqslant f(Z_n)$?

Solution of (i) *by Richard Stong, University of California at Los Angeles.* Observe that $f(G_1 \times G_2) = f(G_1)f(G_2)$ provided $\operatorname{ord}(G_1)$ and $\operatorname{ord}(G_2)$ are relatively prime. Therefore, since a nilpotent group is a direct product of p-groups, it is enough to show the result for p-groups. One easily calculates that

$$f(Z_{p^n}) = \frac{p^{n+1} + p^{-n}}{p+1} > p^{n-1}.$$

[See the following solution of (ii) for details.]

Further if G is any other group of order p^n, then G contains no element of order p^n. Hence no element of G has order exceeding p^{n-1} and $f(G) < p^{n-1}$. Thus if G is a nilpotent group of order n, then $f(G) \leqslant f(Z_n)$ with equality if and only if G is cyclic.

Solution of (ii) *by John H. Lindsey II, Ft. Myers, Florida.* The answer is yes. Let π be a set of primes, let $G_\pi = \{g: g \in G \text{ and } |g| \text{ is divisible only by primes in } \pi\}$, and let $f_\pi(G) = |G_\pi|^{-1} \sum_{g \in G_\pi} \operatorname{ord}(g)$. Let g_p be the p-part of an element g in G, i.e., if $\operatorname{ord}(g) = p^t m$, where $p \nmid m$, then $g_p = g^{am}$, where $1 = am + bp^t$ with $a, b \in \mathbb{Z}$. Let $n = \prod_p n_p$ be the decomposition of n into prime powers. Then

$$\begin{aligned} nf(Z_n) &= \sum_{g \in Z_n} \operatorname{ord}(g) = \sum_{g \in Z_n} \operatorname{ord}\left(\prod_p g_p\right) \\ &= \sum_{g \in Z_n} \prod_p \operatorname{ord}(g_p) = \prod_{p \mid n} \sum_{g_p \in (Z_n)_p} \operatorname{ord}(g_p), \end{aligned}$$

where $(Z_n)_p$ is the p-primary component of Z_n. Thus $f(Z_n) = \prod_{p \mid n} f((Z_n)_p)$. Furthermore,

$$\begin{aligned} f\big((Z_n)_p\big) &= f\big(Z_{n_p}\big) \\ &= n_p^{-1}\big\{1 + p(p-1) + p^2 p(p-1) + \cdots + n_p(n_p/p)(p-1)\big\} \\ &= n_p^{-1}\big\{1 + p(p-1)\big(n_p^2 - 1\big)/\big(p^2 - 1\big)\big\} \\ &= \frac{pn_p}{p+1} + \frac{1}{n_p(p+1)}. \end{aligned}$$

If $n_p > 1$, then

$$f\big(Z_{n_p}\big) = \frac{pn_p}{p+1}\left(1 + \frac{1}{pn_p^2}\right) \leqslant \frac{pn_p}{p+1}\left(1 + \frac{1}{2 \cdot 4}\right) = \frac{9pn_p}{8(p+1)}.$$

Let π_n be the set of odd primes dividing n and let π be any finite set of primes. Put $h_\pi(n) = (2/3)^{|\pi_n - \pi|} f(Z_n)$, so that $h_\pi(n) = \prod_{p|n} h_\pi(n_p)$. Then (ii) follows by taking $\pi = \pi_n \cup \{2\}$ in the following theorem.

THEOREM. *If G is a group of order n, then $f_\pi(G) \leqslant h_\pi(n)$.*

Proof. We may as well replace π by $\pi \cap (\pi_n \cup \{2\})$. Then we use induction first on $|G|$ and next on $|\pi|$. We first show that $h_\pi(n)$ is increasing in the sense that $(h_\pi(m) \leqslant h_\pi(n)$ if $m|n$. In fact, if $m = n/q$ for q a prime, then

$$\frac{h_\pi(m)}{h_\pi(n)} = \frac{h_\pi(n_q/q)}{h_\pi(n_q)}.$$

Furthermore, if $n_q > q$, then

$$\frac{h_\pi(n_q/q)}{h_\pi(n_q)} = \frac{f(Z_{n_q/q})}{f(Z_{n_q})} < \frac{\frac{9}{8}\frac{q(n_q/q)}{q+1}}{\frac{qn_q}{q+1}} = \frac{9}{8q} \leqslant \frac{9}{16}.$$

If $n_q = q$,

$$\frac{h_\pi(n_q/q)}{h_\pi(n_q)} = \frac{1}{h_\pi(q)} \leqslant \frac{1}{\frac{2}{3}f(Z_q)} < \frac{3}{2}\frac{(q+1)}{qq},$$

which is at most

$$\frac{3}{2}\frac{4}{9} = \frac{2}{3}$$

if q is odd. If $q = 2$,

$$\frac{1}{h_\pi(q)} = \frac{1}{f(Z_2)} = \frac{1}{3/2} = \frac{2}{3}.$$

In any event,

$$\frac{h_\pi(n_q/q)}{h_\pi(n_q)} \leqslant \frac{2}{3}.$$

If π is empty, $f_\pi(G) = 1 = h_\pi(1) \leqslant h_\pi(n)$, by monotonicity. For the rest of this paragraph, suppose $\pi = \{2\}$. Suppose $n = n_2$. If G is cyclic, $f_\pi(G) = f(Z_n) = h_\pi(n)$. If G is not cyclic,

$$f_\pi(G) \leqslant \frac{n}{2} < \frac{2n}{2+1} < h_\pi(n).$$

Finally, suppose that p is an odd prime dividing n. By monotonicity

$$h_\pi(n) \geqslant h_\pi(pn_2) = h_\pi(p)h_\pi(n_2) > \frac{2}{3}\frac{p}{p+1}p\frac{2}{2+1}n_2$$

$$\geqslant \frac{2}{3}\frac{3}{4}3\frac{2}{3}n_2 = n_2 \geqslant f_\pi(G).$$

Now, suppose that p is an odd prime in π. Define a *class* to be the set of all π-elements having the same p-part. It suffices to show that the average order over

each class is at most $h_\pi(n)$. Suppose that the π-element $g = xy$ has p-part x. Then $\mathrm{ord}(g) = \mathrm{ord}(x) \cdot \mathrm{ord}(y)$ and y is a $(\pi - \{p\})$-element in $C_G(x) = C(x)$. Hence $y \mapsto y\langle x\rangle$ is 1-1 ord()-preserving correspondence between such y and the $(\pi - \{p\})$-elements of $H = C(x)/\langle x\rangle$. Thus the average of ord() over this class is $\mathrm{ord}(x) f_{\pi-\{p\}}(H)$. Suppose $\mathrm{ord}(x) < n_p$. Using the induction hypothesis (since $|\pi - \{p\}| < |\pi|$) and monotonicity,

$$\begin{aligned}
\mathrm{ord}(x) f_{\pi-\{p\}}(H) &\leqslant \mathrm{ord}(x) h_{\pi-\{p\}}(|H|) \leqslant \mathrm{ord}(x) h_{\pi-\{p\}}(|G|/\mathrm{ord}(x)) \\
&= \mathrm{ord}(x) h_{\pi-\{p\}}(n_p/\mathrm{ord}(x)) h_\pi(n/n_p) \\
&= \mathrm{ord}(x) \frac{2}{3} f(Z_{n_p/\mathrm{ord}(x)}) h_\pi(n/n_p) \\
&\leqslant \mathrm{ord}(x) \frac{2}{3} \frac{9}{8} \frac{p}{p+1} \frac{n_p}{\mathrm{ord}(x)} h_\pi(n/n_p) < \frac{p}{p+1} n_p h_\pi(n/n_p) \\
&< h_\pi(n_p) h_\pi(n/n_p) = h_\pi(n).
\end{aligned}$$

Therefore we are through unless there exists x of order n_p. Suppose $C(x) \subsetneqq G$. Then for some prime $q \neq p$, $|C(x)| \,|\, |G|/q$. Using the induction hypothesis and monotonicity, we see that the average of ord() over this class is

$$\begin{aligned}
n_p f_{\pi-\{p\}}(H) &\leqslant n_p h_{\pi-\{p\}}(|H|) = n_p h_\pi(|H|) \leqslant n_p h_\pi(|G|/(qn_p)) \\
&= n_p h_\pi(n/n_p) \frac{h_\pi(n_q/q)}{h_\pi(n_q)} \leqslant n_p h_\pi(n/n_p) \frac{2}{3} \leqslant \frac{p}{p+1} n_p h_\pi(n/n_p) \\
&\leqslant h_\pi(n_p) h_\pi(n/n_p) = h_\pi(n).
\end{aligned}$$

The only case where our class-average proof breaks down is when $\mathrm{ord}(x) = n_p$ and $C(x) = G$. Then, by Schur-Zassenhaus, $G \cong \langle x\rangle \times H$ for some p'-group H. By a calculation similar to that for $nf(Z_n)$ and by the induction hypothesis applied to H, we find that

$$\begin{aligned}
f_\pi(G) &= f(Z_{\mathrm{ord}(x)}) f_\pi(H) = h_\pi(\mathrm{ord}(x)) f_\pi(H) \\
&\leqslant h_\pi(\mathrm{ord}(x)) h_\pi(|H|) = h_\pi(|G|).
\end{aligned}$$

(We do not really need Schur-Zassenhaus, just the bijection on G_π: $g \mapsto (g_p, g_{p'})$ where independently g_p runs over $\langle x\rangle$ and $g_{p'}$ runs over the p'-elements of G_π which are in 1-1 correspondence with $(G/\langle x\rangle)_\pi$.) This completes the proof.

No other solutions were received for Part (ii). Part (i) was solved also by David Callan, O. P. Lossers, the late David Richman, John Henry Steelman, and the proposer.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the Monthly presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

MONTHLY UNSOLVED PROBLEMS, 1969–1991

RICHARD K. GUY

This is the eleventh of the two-yearly articles that have appeared since the inception of this section of the *Monthly*. It is written in June, but, as I prepare to serve under my sixth editor, I make a New Year's resolution to keep the articles flowing more regularly. In part this depends on you to keep me supplied with suitable problems, and these are not easy to find [1971, 1113–1114; 1973, 1120–1121; 1981, 755–756; 1983, 122].

References in brackets are to year and page numbers of this *Monthly*, while dates in parentheses refer to publications listed at the end, and other items are labelled (tbp) if they are likely to be published formally, or as written communications (wrc) if publication plans are not presently known. Dates and pages in brackets are also appended to items in the bibliography indicating where the problem originally appeared in the *Monthly*.

We mention here some other collections of unsolved problems. Croft, Falconer and Guy have recently published *Unsolved Problems in Geometry* (1991) [*UPIG*] and I have edited collections of unsolved problems in number theory (1989, 1990), as well as the informal collections issued in connection with the annual Western Number Theory conferences, and I hope that the second edition of the book [*UPINT*] will be available by the time this article is being read. A collection of 37 unsolved problems in combinatorial games appears on pages 183–189 of Guy (1991).

It would seem that Joseph Gallian [1990, 133], with his survey paper (1989) and two other papers (1987, 1988) has become keeper of the records for the Kotzig-Ringel tree-labelling problem [1969, 1128]. My bibliography, which needs to be put into electronic form, contains 185 items.

In discussing the famous Sylvester problem and its relation to the magic configuration problem of Murty [1971, 1000] I misquoted [1989, 903–905] a conjecture of Smyth. It should have read "¿If a configuration contains more than five special points, then all special points are collinear?" A counterexample, for which Smyth did not have the source, is provided (for $k \geqslant 2$) by the set T of the vertices and the centre, c, of a regular $(4k + 2)$-gon, together with the set $S = \{s_1, \ldots, s_{2k+1}\}$ of points at infinity on the sides (and diagonals) of the polygon. The special points are $c \cup S$ which are $2k + 2$ in number and not all collinear. Smyth's other conjecture: "¿The only nontrivial configuration in which the number of special points exceeds the number of ordinary points is the failed Fano plane?" remains uncontradicted.

Füredi (1991) has extended to $w < \sqrt{3}$ the solution of Fejes Tóth's [1971, 528] problem asking for the densest packing of points (resp. circles of unit diameter) in

an infinite strip of width w (resp. $w + 1$), no two points being distance less than 1 apart.

For those of us who have struggled with crossing-number problems [1973, 52] it is some comfort that Bienstock (wrc) has shown that some of them are provably hard. He and Dean (wrc) have also described, for each $k \geqslant 4$, a class of graphs with crossing number k, but unbounded *rectilinear* crossing number. An interesting breakthrough has been made by Douglas Woodall (tbp) towards verifying what may be called Zarankiewicz's conjecture, that the (minimal) crossing number of $K_{m,n}$, the complete bipartite graph on $m + n$ vertices, is equal to the known upper bound

$$\nu(K_{m,n}) \leqslant \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor$$

by showing its truth in the cases $\nu(K_{7,7}) = 81$ and $\nu(K_{7,9}) = 144$. A counting argument then implies that the formula holds for $(m, n) = (7, 8), (8, 8), (7, 10)$ and $(8, 10)$. Unfortunately, the argument used by Kleitman (1970) to deduce the case $(5, n)$ from his result for $(5, 5)$ doesn't seem to work for $(7, n)$, so the smallest unknown values are for $(7, 11)$ and $(9, 9)$. The counting argument shows that $\nu(K_{7,11}) \geqslant 220$ and a parity argument shows that the value is odd, so it is 221, 223 or 225. Similarly, $\nu(K_{9,9})$ is 248, 250, 252, 254 or 256. Which?

Gerver (wrc) has produced a "sofa" [1976, 188] of area 2.2195... which can be moved round a right-angled corridor of width 1, which beats the earlier record of 2.215649... [1977, 811 and see *UPIG*, **G5**] and which he conjectures to be best possible.

Students of Conway's game of Sylver Coinage [1976, 634] should consult the tables on pp. 171–172 of Richard Nowakowski's chapter in Guy (1991), where the results of George Sicherman's extensive calculations are incorporated. Nowakowski is also the natural person to appoint as repository for any additional information about this game.

We announced [1987, 963; 1989, 905] that Tunnell had effectively solved the "congruent number" problem [1980, 43]: Monsky's (1990) paper on this topic has since appeared. Table 1 [1981, 759] is now complete and will appear as Table 7 in the second edition of *UPINT*.

Butler (1991) answers all three of Piranian's [1980, 555] questions—see (a), (b), (c) at foot of [1991, 139].

The interesting sidelight thrown by Lagarias (1990) on the notorious $3x + 1$ problem [1983, 35; 1985, 3–23] has now appeared. He is the natural repository for information on this problem, which he writes that he is not done with. He plans a joint paper on the heuristics for the total stopping time. There are two distinct probability models: a random walk based on mod 2 behavior, and a branching process based on mod 3 behavior (going backwards). Both models can be solved exactly in principle and predict a maximum excursion of $(c_0 + o(1)) \ln N$ as $N \to \infty$, where $c_0 = 41.1\ldots$ is the same constant for both models.

At a recent Western Number Theory conference Basil Gordon asked if it is true that for every positive integer n there is a one-one map L of $\{1, 2, \ldots, n\}$ onto $\{0, 1, \ldots, n-1\}$ such that $L(ab) = L(a) + L(b)$ whenever a, b and ab are all in $\{1, 2, \ldots, n\}$. It is easy to confuse this problem with that of Forcade, Lamoreaux and Pollington [1986, 119; see also 1989, 905] where the mapping is to be onto a **group** (called an FLP group below). Gordon's problem is a special case, so the counterexamples found by Forcade and Pollington (1990) for their problem also

serve as counterexamples for Gordon's. Contrast the two examples for $n = 10$:

	1	2	3	4	5	6	7	8	9	10
L_P	0	1	8	2	4	9	7	3	6	5
L_G	0	1	4	2	6	5	9	3	8	7

The former is obtained by using 2 as a primitive root of $n + 1 = 11$ and the entries, which are exponents (logarithms) are the elements of the additive group of residue classes mode 10. However $L(3) + L(3) = 8 + 8 \neq 6 = L(3 \times 3)$ unless the addition is so interpreted. We can satisfy Gordon's requirements by rearranging the entries as in the second example. Forcade and Pollington gave the counterexamples 195 and 255 to their own problem and 195 is the least such. Perhaps $n = 105$ (but not $n = 35$) is the smallest counterexample to Gordon's problem; John Selfridge has done a backtrack on this by hand, but it should be confirmed by machine. Chandler (1988) has shown that every odd order FLP group is commutative.

Tomaszewski [1986, 280] considered n real numbers $a_1, \ldots, a_n$ satisfying $\sum_{i=1}^{n} a_i^2 = 1$ and asked if, of the 2^n sums of the form $\sum \pm a_i$, it is possible that there are more with $|\sum \pm a_i| > 1$ than there are with $|\sum \pm a_i| \leqslant 1$. Ron Holzman (wrc) proves that the proportion with smaller sum is at least $63/169 > 0.37278$.

It seems to be very hard to determine whether or not there is a point in the plane of a unit square which is at a rational distance from each of its corners [1988, 33 and see *UPINT*, **D19**], but Bremner and Guy (1989 and tbp) have in some sense found all tilings of the square with four rational triangles.

In [1988, 35] we also asked for integer solutions of the simultaneous equations

$$x^2 + y^2 + z^2 = a^2 + b^2 + c^2 \qquad x^2y^2z^2 = a^2b^2c^2$$

and later [1989, 907] gave some history. Because of our backlog, John B. Kelly's neat general solution (1991) has taken longer to appear than expected.

Murray Klamkin [1988, 845] conjectured that if a curve subtended a constant angle from the points of each of two concentric circles, then it was itself a circle. Johannes Nitsche (1990) proved this, as did Gerd Baron, and suggested that it might be of interest to consider similar questions in higher dimensions.

A certain amount of correspondence concerning Shyam Sunder Gupta's [1989, 425] quest for decimal numbers whose sum and difference with their reversals are both squares, has yielded no examples other than those reported in [1989, 908]. Are there infinitely many non-palindromic examples? Are there any odd, and therefore necessarily non-palindromic, examples, besides $65 \pm 56 = 11^2, 3^2$?

Martin Aigner writes concerning Bill Sands's guessing game [1990, 314] that if an unknown element x from $[1, \ldots, n]$ is to be deduced from (truthful) answers of form "$x = k$" or "$0 < |x - k| \leqslant d$" or "$x - k| > d$" to questions "Does $x = k$?" and if $L_d(n)$ is the number of answers in the worst case, and $2^{h-2} < d \leqslant 2^{h-1}$, then, for $n \geqslant 2d(h + 1) + 2h$,

$$L_d(n) = \begin{cases} \left\lceil \dfrac{n - 2^h - h}{2d + 1} \right\rceil + h & \text{for } n \neq 2^{h-2} + 1, \quad n \geqslant 4 \\ \left\lceil \dfrac{n - 2^h - h - 1}{2d + 1} \right\rceil + h & \text{for } n = 2^{h-2} + 1, \quad n \geqslant 5 \end{cases}$$

$$L_3(n) = \left\lceil \frac{n - 6}{7} \right\rceil + 2, \qquad L_2(n) = \left\lceil \frac{n - 2}{5} \right\rceil + 1, \qquad L_1(n) = \left\lceil \frac{n}{3} \right\rceil$$

with larger values for $L_d(n)$ if $n < 2d(h + 1) + 2h$. Howard Cary Morris and others also obtained partial results. Variations in which there is, or may be, an untruthful answer, may be worth considering.

A. Zulauf of the University of Waikato, New Zealand, has proved that the first player can win the game of Barranca [1990, 315] by first choosing 2 and eventually achieving a score of 224, 216 or 210 (it is easy for each player to prevent the opponent from obtaining a perfect score of 220). The complete analysis is too extensive to publish here.

REFERENCES

Daniel Bienstock, Some provably hard crossing number problems (preprint) [1973, 52].

Dan Bienstock and Nate Dean, Bounds for rectilinear crossing numbers (preprint) [1973, 52].

Andrew Bremner and Richard K. Guy, Delta-lambda configurations in tiling the square, *J. Number Theory*, 32 (1989) 263–280; *MR* 90g: 11031 [1988, 33].

Andrew Bremner and Richard K. Guy, Nu-configurations in tiling the square, *Math. Comput.*, 58(1992) (to appear) [1988, 33].

James P. Butler, The perimeter of a rose, this *Monthly*, 98(1991) 139–143 [1980, 555].

K. A. Chandler, Groups formed by redefining multiplication, *Canad. Math. Bull.*, 31(1988) 419–423; *MR* 89m:20021 [1986, 119].

Hallard T. Croft, Kenneth J. Falconer and Richard K. Guy, Unsolved Problems in Intuitive Mathematics, Vol. II, Unsolved Problems in Geometry [*UPIG*], Springer-Verlag, 1991.

R. W. Forcade and A. D. Pollington, What is special about 195? Groups, nth power maps and a problem of Graham, in R. A. Mollin (editor) Number Theory, *Proc. 1st Conf. Canad. Number Theory Assoc., Banff, 1988,* de Gruyter, 1990, pp. 147–155 [1986, 119].

R. Frucht and J. A. Gallian, Labeling prisms, *Ars Combin.*, 26(1988) 69–82; *MR* 90a:05178 [1969, 1128].

Zoltán Füredi, The densest packing of equal circles into a parallel strip, *Discrete Comput. Geom.*, 6(1991) 95–106 [1971, 528].

Joseph A. Gallian, Labeling prisms and prism related graphs, in Proc. 18th SE Internat. Conf. Combin. Graph Theory, Comput., Boca Raton, *Cong. Numerantium*, 59(1987) 89–100; *MR* 89d:05162 [1969, 1128].

Joseph A. Gallian, A survey—recent results, conjectures and open problems in labeling graphs, *J. Graph Theory*, 13(1989) 491–504; *MR* 90k:05132 [1969, 1128].

Joseph L. Gerver, On moving a sofa around a corner (preprint) [1976, 188].

Richard K. Guy (editor), Unsolved Problems, in R. A. Mollin (editor) Number Theory and Applications, *Proc. NATO ASI, Banff 1988*, Kluwer, Dordrecht, 1989, 593–602.

Richard K. Guy, Canadian Number Theory Association unsolved problems, 1988, in R. A. Mollin (editor), Number Theory, *Proc. 1st Conf. Canad. Number Theory Assoc., Banff, 1988,* de Gruyter, 1990, pp. 193–206.

Richard K. Guy, Unsolved Problems in Intuitive Mathematics, Vol. I, Unsolved Problems in Number Theory [*UPINT*], 2nd edition, Springer-Verlag, 1992.

Richard K. Guy (editor), Combinatorial games, *Proc. Symp. Appl. Math.*, 43, Amer. Math. Soc., Providence, 1991 [1976, 634].

Jeffery C. Lagarias, The set of rational cycles for the $3x + 1$ problem, *Acta Arith.*, 56(1990) 33–53 [1983, 35].

Ron Holzman, On the product of sign vectors and unit vectors (preprint) [1986, 280].

John B. Kelly, Two equal sums of three squares with equal products, this *Monthly*, 98(1991) 527–529 [1988, 35].

Daniel J. Kleitman, The crossing number of $K_{5,n}$, *J. Combin. Theory*, 9(1970) 315–323; *MR* 43 #6123 [1973, 52].

Paul Monsky, Mock Heegner points and congruent numbers, *Math. Z.*, 204(1990) 45–67; *MR* 91e:11059 [1980, 43].

Johannes C. C. Nitsche, Isoptic characterization of a circle (proof of a conjecture of M.S. Klamkin), this *Monthly*, 97(1990)45–47; *MR* 90j:51029 [1988, 845].

Douglas R. Woodall (tbp), Cyclic-order graphs and Zarankiewicz's crossing-number conjecture [1973, 52].

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Recent Advances in Geometric Inequalities. By D. S. Mitrinović, J. E. Pečarić and V. Volonec, Kluwer Academic Publishers, Dordrecht/Boston/London, 1989, xix + 710 pp.

DAN PEDOE
Professor Emeritus, School of Mathematics, University of Minnesota, Minneapolis, MN 55455

A book of a mere 151 pages entitled *Geometric Inequalities*, by O. Bottema, R. Ž. Djordjević, R. R. Janić, D. S. Mitrinović, and P. M. Vasić, published in 1969, contained about 400 inequalities and listed 225 authors, and aroused a great deal of interest. The giant of a book under review derives from this sturdy infant, contains 750 names of those who have worked on inequalities up to the year 1986, and lists thousands of inequalities. It also describes conjectures, unsolved problems, and research dating back to the 19th century. Publications in Chinese, Japanese, Serbo-Croatian, Bulgarian, Rumanian, Hungarian and Dutch are reviewed, besides, of course, those in English, French, German and Russian. The scope of this work dazzles the imagination. The series editor, Michiel Hazewinkel, ends the first Preface in the book with a challenge:

> Well, if you know of a better 'ole, go to it!
> Bruce Bairnsfather

The reviewer was a child in WWI, and remembers the cartoon by Bruce Bairnsfather, grimly designed to reassure the British public that trench warfare could be made bearable. The cartoon showed a rubicund Cockney smoking a pipe, cooking sausages over a paraffin stove, quaffing a beer, in a comfortable-looking trench illuminated by candles. The series editor also gives a short quotation from a G. K. Chesterton novel, another from *The Chinese Maze Murders* by a R. van Gulik, states the title of a famous paper by Eugene Wigner, and ends his quotations with one from William Blake and one from Joseph Louis Lagrange.

In other words the series editor makes it plain that the world is his oyster (if he is a she, I apologize), and will not have his subject deprecated. He points out that geometric inequalities have important applications both to geometry and to areas well beyond the traditional areas of geometry, such as the theory of complex functions and the calculus of variations, to the theory of embedding theorems for function spaces, and more generally in providing *a priori* estimates in several areas, such as differential equations.

In any triangle with sides a, b and c, we have the inequalities

$$b + c > a, \qquad c + a > b, \qquad a + b > c.$$

This fundamental set of inequalities, naturally called *triangle inequalities*, appears in Euclid, but certainly precede Euclid, being known to any donkey seeking a bundle of hay for rapid consumption: the shortest distance between any two points

is the straight line segment joining them. Chapter I of the book under review is concerned with the existence of a triangle satisfying more complex conditions, beginning with inequalities involving the radius of the circumcircle, the radius of the incircle, and the semi-perimeter. There are 48 references given at the end of this chapter, which sets the tone for the whole book, since essential historical descriptions are given in the text, links between theorems are traced, and short proofs are given where appropriate. Möbius appears in the references, with an 1852 date, and, to come to modern times, Paul Erdös appears in a joint note with Murray Klamkin.

Chapter II deals with "Duality Between Geometric Inequalities," Chapter III with "Homogeneous Symmetric Polynomial Geometric Inequalities," Chapter IV with "Duality Between Different Triangle Inequalities and Triangle Inequalities with (R, r, s),"..., no aspect of geometric and related inequalities is neglected.

In 1941, waiting for the inevitable daylight air-raid on Southampton, I was doodling with pencil and paper, and wondered whether the orthogonal projection of a given triangle into an equilateral triangle could be extended to the orthogonal projection of a given triangle into a triangle of given shape. In proving that this could always be done, I discovered a two-triangle inequality for triangles with respective sides a, b, c and a', b', c' and areas Δ and Δ':

$$a^2(-a'^2 + b'^2 + c'^2) + b^2(a'^2 - b'^2 + c'^2) + c^2(a'^2 + b'^2 - c'^2) \geqq 16\Delta\Delta',$$

with equality iff the two triangles are similar.

I thought that this was an interesting inequality, and it reminded me of the general inequality which connects the *mixed area* of two convex domains with their respective areas. I tried to have my inequality published, but some people at Oxford were disturbed by the term *orthogonal projection* and rejected publication. The late Beniamino Segre, a great geometer, a refugee from Italy, was in a depressed mood, and labelled it "Mathematical Tripos" stuff (which was curious, because the Tripos in those days contained a lot of fine geometry). The inequality might have been forgotten if I had not discovered that outside Britain well-known geometers such as Hadwiger had been very interested in mere single-triangle inequalities. I used another geometrical method, not involving orthogonal projection, for proving my inequality, and this was accepted as a Research Note for the Cambridge Philosophical Society Proceedings.

J. Neuberg, a 19th-century geometer, had discovered the inequality, and remarked that this was evidently an equality if the two triangles are similar, but had not proved that equality implied the similarity of the two triangles. For this reason the theorem was labelled the Neuberg-Pedoe inequality.

In recent times this inequality has attracted a fair amount of attention. Two young mainland-Chinese geometers have generalized the inequality to n-simplexes in Euclidean n-space, incidentally showing that so-called college algebra should not be neglected. The book under review has a section of 15 pages on *The Neuberg-Pedoe and the Oppenheim Inequalities*, with 38 references.

The last chapter of 100 pages is on inequalities in E^n, with 129 titles listed at the end of the chapter. Before that, inequalities are considered for quadrilaterals, polygons, circles, and there are *Some Particular Inequalities* which, of course,

consider isoperimetric inequalities. Many of the references go back to the 19th century, but one special one caught my attention:

S. A. J. Lhuilier: *De relatione mutua capatitatis et terminorum figurarum, geometrice considerata: seu de maximis et minimis*, Varsaviae, 1782.

My 1941 doodling with the orthogonal projection of a given triangle into one of a given shape is, of course, Lhuilier's theorem, that a 3-edged prism can be cut by a suitable plane in a triangle of any given shape. Priority is acknowledged in "Homogeneous Coordinates and the Lhuilier Theorem," Crux Mathematicorum, vol. 9, 1983, pp. 160–165, which was also a response to Dieudonné's defiant "Who ever uses homogeneous coordinates?"

Lest it be thought that this remarkable book is too highbrow for the average reader, the authors stress that their book is intended for a wide circle of readers, for students attending high schools, colleges and universities, as well as for their professors. To show the great variety among those listed in the index of names in this book, I list those known to me, either personally, or by correspondence.

Sir Alexander Oppenheim appointed me to the University of Singapore. He was knighted for his services to Commonwealth education, and, in particular, for running a prison-camp university when Singapore was overrun by the Japanese army. Hidetosi Fukagawa is a high-school teacher in Japan, in his late forties. I have collaborated with him over the past six years in writing *Japanese Temple Geometry Problems* (reviewed in this journal, April, 1991), the product of geometry enthusiasts in 18th- and 19th-century Japan who painted their theorems on wooden tablets and hung them in the precincts of shrines and temples. These enthusiasts came from all social classes in Japan. George Tsintsifas is the principal of a private high school in Saloniki, Greece. Walther Janous teaches at an Ursuline convent in Innsbrück, Austria.

Jordan Tabov, whose wife is also a mathematician, runs mathematics competitions, in which he distinguished himself as a student, in Sofia, Bulgaria. L. Yang and J. Zh. Zhang are the two young mainland-Chinese mentioned earlier. They could be called the alpha and omega of geometry in China, producing remarkable new theorems which can be proved by the use of computers and Hilbert's theorems on polynomials. They first appeared in *Crux Mathematicorum*, demonstrating the amazing constructions possible using only a "rusty" compass (one of fixed radius). Geometry in mainland China seems to have survived the "cultural" revolution. Murray Klamkin has worked for the Ford Motor Company. He is once again a university professor and a great collector and solver of problems. J. F. Rigby is an English academic, and one of England's leading geometers. Donald Coxeter has lived in Canada for many years, and can be called the G.O.M. of geometry there. Roland Eddy hails from St. John's University, Newfoundland, and was encouraged by his professor, W. J. Blundon, prominently featured in this book, to study geometry, and has justified his professor's insight. The great Jordi Dou is a retired architect, and lives in Barcelona. Geometry is his natural element. Bernhard Neumann has lived in Australia for many years, and is a very distinguished mathematician. The collected papers of Neumann and his late wife Hannah have just been published by the same press which published my book written with Fukagawa. The late I. M. Jaglom is well known because of the translations from the Russian of his books on geometry. T. J. Fletcher was an Inspector of Schools for Her Majesty's Government some years ago, wrote excel-

lent textbooks and pioneered in the making of films on geometry. S. Iwata is the editor of a remarkable book, an *Encyclopaedia of Geometry*, in Japanese, and one of the circle of devotees who meet to discuss Japanese Temple Geometry Problems. My collaborator, Hidetosi Fukagawa, owes much to Prof. Iwata's encouragement over the years. V. C. Linis and Fred Maskell were strong supporters of Léo Sauvé, the great editor of *Crux*, in its early days. They are all, alas, departed, but not forgotten. L. J. Mordell is still well known for his work in number theory, and was professor in Manchester for many years.

Dmitri Mavlo wrote charming letters from Moscow in excellent English to both Léo Sauvé and myself, with photographs, inviting us for a tour of the Soviet Union, and suggesting possible problems for *Crux*. He no longer writes, and there is a hint in his last letter, before the Gorbachev era, that he had been in trouble with the regime. Leon Bankoff is a dentist in Los Angeles and has dramatized the beauties of geometry for many years. Bob Osserman is an excellent geometer whose paper "Curvature in the Eighties" was one of six in the special geometry issue of the MONTHLY, October, 1990. Van der Waerden is probably still remembered for his magnificent books on modern algebra, originally published in German, and began the formidable task of putting algebraic geometry on a sound footing, with a series of papers in Mathematische Annalen. Paul Erdös, of course, is known to the whole world, and is unique in many ways.

Last, but not least, J. C. (Chris.) Fisher, D. Ruoff and J. Shilleto are of the new generation of active geometers in North America who enjoy working and publishing together.

It is clear that there will be another edition of the great work under review, perhaps in ten years from now. The list of names is bound to grow exponentially. The authors welcome comments and corrections, and are obviously sincere in this. Having written two books with a coauthor, I cannot even imagine the logistics of cooperation between three authors which produced this book. Perhaps, as is the case with the making of some movies, there will be a book which will explain how the book was produced. In the meantime, I feel that this may well be the most inspirational book of the decade which has just passed, setting a pattern for cooperation in the joys of mathematical discovery which few have ever dreamed of. A world congress on geometric inequalities, which is very possible, would be a major cultural event of our time. For the immediate future, however, this book should be (possibly chained!) in every university and college library, and, yes, in the library of every school which is intent on improving its mathematics teaching.

IN APPRECIATION

At this season, the Monthly expresses its appreciation to the following people for their help in refereeing submitted manuscripts during the past year. Without the help, the caring and the dedication of our referees, the Monthly could not function.

Janos D. Aczel, A. Aksoy, Felix R. Albrecht, Stephanie B. Alexander, J. Ralph Alexander, Bruce A. Anderson, George E. Andrews, Tom M. Apostol, Kenneth I. Appel, Michael Artin, Mark S. Ashbaugh, Richard A. Askey, J.C. Ault, Frank Baginski, Craig Bailey, Steven B. Bank, Edward Barbeau, Helénè Barcelo, Michael Barr, Robert Bartle, Sybilla Beckmann, Lowell W. Beineke, Katalin Bencsáth, A. Benjamin, Grahame Bennett, D. Bentley, I. David Berg, Bruce C. Berndt, Amer Beslagić, Patrick Billingsley, Richard L. Bishop, T. Bisztriczky, Harold B. Boas, Joseph Bonin, Peter B. Borwein, Nigel Boston, Bart Braden, David M. Bradley, Fred Brauer, William E. Briggs, John D. Brillhart, Duane M. Broline, Richard Brualdi, R.H. Bruck, Andrew W. Bruckner, Victor Bryant, Donald L. Burkholder, Robert Burton, Lynne Butler,

C. David Callan, Bille C. Carlson, Robert W. Carroll, Frank S. Cater, Alfred S. Cavaretta, Gulbank D. Chakerian, J.H.H. Chalk, Laura Chihadra, Erhan Cinlar, Francis H. Clarke, Matthew Cook, K. Cooke, Robert F. Craggs, E. Cumberbatch, Frederic Cunningham, Thomas W. Cusick, John P. D'Angelo, Everett C. Dade, Mahlon M. Day, J. Day, N.G. DeBruijn, Larry L. Dornhoff, Underwood Dudley, Paul Edelm, Gerald Edgar, Omer Egecioglu, R. Elderkin, James F. Epperson, Paul Erdös, Ronald J. Evans, Charles L. Vanden Eynden, Doug Faires, Kenneth J. Falconer, Philip Feinsilver, Michael A. Filaseta, James Fill, Nathan J. Fine, Dan Flath, George K. Francis, Jane Friedman, Stephen M. Gagola, David Gale, Joseph A. Gallian, Fred Galvin, Ira M. Gessel, Richard A. Gibbs, John Gimbel, Clark Givens, Solomon Golomb, Michael Golomb, Irving J. Good, Adolf W. Goodman, Victor Goodman, J. Grabiner, S. Grabiner, Ronald L. Graham, Sidney W. Graham, Alfred Gray, Daniel R. Grayson, R.J. Gregorac, Ralph Grimaldi, William Gustafson,

Peter Hagis, Heine Halberstam, Mary-Elizabeth Hamstrom, Frank Harary, David Harbater, Lester L. Helms, Joe Hemmeter, Douglas A. Hensley, Adolf Hildebrand, Aimo Hinkkanen, Kevin Hockett, Robert V. Hogg, Thomas W. Hungerford, I. Martin Isaacs, John R. Isbell, Mourad E.H. Ismail, Nathan Jacobson, Gerald J. Janusz, Richard P. Jerrard, Carl G. Jockusch, Charles R. Johnson, Norman Johnson, James P. Jones, D. Kalman, Gerald Kaminski, Kenneth Kaminsky, Robert P. Kaufman, Nicholas D. Kazarinoff, Clark H. Kimberling, Allan Kirch, Victor L. Klee, Daniel J. Kleitman, Neal Koblitz, Fred Kochman, Joseph Konhauser, R. Kopperman, Tsit-Yuen Lam, Ray G. Langebartel, Charles F. Laywine, Felix Lazebnik, Arnold Lebow, Carl Lee, Myrna Lee, Derrick H. Lehmer, Joseph Lehmer, Emma Lehmer, W. Lucas, Carsten Lund, Frederick W. Luttmann, Russell D. Lyons,

Hosam Mahmoud, Daniel Maki, Marvin Marcus, David E. Marker, M. Martelli, R. Daniel Mauldin, Jim Mauldon, Leon R. McCulloch, Lynn McLinden, George F. McNulty, Russel Merris, P. Meyer, Joseph B. Miles, Frank B. Miles, Hugh L. Montgomery, John C. Morgan, David E. Muller, D. Mullins, Melvyn B. Nathanson, Morris Newman, Karl K. Norton, Alec Norton, Andrew M. Odlyzko, J.

Marshall Osborn, Robert Osserman, James Owings, Donald Passman, N. Tenney Peck, G.W. Peck, Rhodes Peele, Richard E. Pfeifer, Richard S. Pierce, A. Pixley, Carl Pomerance, Horacio A. Porta, Stephen V. Portnoy, Jim Propp, George B. Purdy, Bruce A. Reznick, Norman Richert, David R. Richman, Bruce Richter, Joseph B. Roberts, Wayne Roberts, Derek J.S. Robinson, C. Ambrose Rogers, Kenneth A. Ross, Bruce L. Rothschild, Joseph J. Rotman, David E. Rowe, Ranjan Roy, Lee A. Rubel, Walter Rudin,

Donald G. Saari, Hans Sagan, Bill Sands, Jonathan Schaer, Edward Scheinerman, James H. Scherl, Andrzej Schinzel, Thomas Schmidt, Frank W. Schmidt, Eric Schmutz, Lowell Schoenfeld, Paul E. Schupp, David B. Secrest, John L. Selfridge, M. Senechal, Richard D. Shafer, Jeffrey O. Shallit, Daniel Shanks, Daniel B. Shapiro, Louis Shapiro, S. Shariari, Donald R. Sherbert, Lance W. Small, J. Laurie Snell, William M. Snyder, Lawrence Somer, H.M. Srivastava, Richard P. Stanley, Ralph G. Stanton, Dennis Stanton, John Steinig, William Stenger, Dorothy M. Stone, Arthur H. Stone, Gilbert Strang, Robert Strichartz, Keith Stroyan, Lajos F. Takacs, Gaisi Takeuti, F. Tall, M. Townsend, Edward C. Turner, Daniel Ullman, Stephen V. Ullman, Daniel J. Velleman, Samuel S. Wagstaff, Walter Wallis, John H. Walter, Edward T.H. Wang, Lawrence C. Washington, William C. Waterhouse, William A. Webb, Paul M. Weichsel, Elliott C. Weinberg, Howard J. Weiner, Gregory P. Wene, Alan Wessel, John E. Wetzel, Robert A. Wijsman, Jet Wimp, Andrew Woldar, Katherine Yerion, Peter Yff, William R. Zame, Peter Zvengrowski,

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S(15-17), L.** *Introduction to Classical Mathematics I: From the Quadratic Reciprocity Law to the Uniformization Theorem.* Helmut Koch. Transl: John Stillwell. Math. & Its Applic., V. 70. Kluwer Academic, 1991, xvii + 453 pp, $149. [ISBN: 0-7923-1231-7] A unique volume providing clear, concise highlights of nineteenth-century mathematics (from Gauss to Riemann) in historical order but with consistent modern notation and proofs. Assuming two years of (European-level) university mathematics—which is summarized in twenty pages of appendices—the text covers virtually all of classical mathematics with refreshing economy. Extensive integrated historical notes provide narrative and historically-rooted rationale. No exercises. LAS

General, P. *Frontiers in Pure and Applied Mathematics.* Ed: Robert Dautray. North-Holland (US Distr: Elsevier Science), 1991, 286 pp, $108.50. [ISBN: 0-444-88133-6] Nineteen research papers presented at a Paris symposium held to celebrate the 60th birthday of Jacques-Louis Lions. Topics touch on a wide variety of structures and themes (e.g., nuclear energy, Hamiltonian systems, turbulence, immunology) reflecting the diverse influence of Lions' work. Numerous photos from the symposium introduce the volume. LAS

General, S(15-18), P, L. *Symmetry of Discrete Mathematical Structures and Their Symmetry Groups: A Collection of Essays.* Eds: K.H. Hofmann, R. Wille. Res. & Expos. in Math., V. 15. Heldermann Verlag, 1991, vi + 168 pp, $68 (P). [ISBN: 3-88538-215-6] Five essays, by distinguished mathematicians, delivered at a large symposium on the role of symmetry in the arts, architecture, literature, music, the natural science, and in every-day life. The contributions of Bokowski, Schulte, and Wills exemplify the function of symmetry in modern mathematical research; those of Dress, Wille, and Hofmann (each in German) are scholarly essays on the concept of symmetry in mathematics, but are general in nature and require no training in mathematics. LCL

Elementary, T(13), S. *Practical Problems in Mathematics for Heating and Cooling Technicians, Second Edition.* Russell Devore. Delmar, 1991, viii + 248 pp, $10.95 (P). [ISBN: 0-8273-4062-1] Arithmetic, measurement, areas, arcs, volumes, etc., applied to pipe and duct geometry, heat load calculations, estimates and bills, and other common plumbing tasks. Workbook style, with space for answers to be written in. LAS

Mathematics Appreciation, T(13-14). *Mathematical Modeling for the Marketplace: Applying Graph Theory in Liberal Arts and Social and Management Sciences.* Helen Christensen. Kendall Hunt, 1988, vii + 167 pp, $18.95 (P). [ISBN: 0-8403-4661-1] Graphs and digraphs provide tools and techniques to model and solve problems from a variety of liberal arts and social science disciplines. Intended for a "liberal arts mathematics" course; accessible with minimal mathematical background, yet stimulating for better-prepared students. Includes examples of student projects, and chapter on inductive reasoning. MW

Education, C, P. *How to Use the Spreadsheet as a Tool in the Secondary School Mathematics Classroom.* William J. Masalski. NCTM, 1990, 76 pp, $16.50 (P); 2 diskettes included. [ISBN: 0-87353-303-8] Thirty spreadsheet activities including magic squares, root-finding by successive approximation and Newton's method, and probability simulations. Computer data files provided for Apple Works and Better Working Spreadsheets on Apple IIe's, but book contains enough detail to allow translation to other spreadsheets. Both problem-solving and electronic blackboard activities. Useful for methods courses and teacher-enhancement projects. MW

Education, T, P. *Teaching and Learning Computer Programming: Multiple Research Perspectives.* Ed: Richard E. Mayer. Lawrence Erlbaum Assoc, 1988, xvi + 320 pp, $49.95. [ISBN: 0-8058-0073-5] Reports of recent theory-based research studies on Logo, BASIC, and Pascal programming. Although early research did not support initial strong claims of discovery-learned programming as a vehicle for learning problem solving skills, these studies suggest some transfer is achieved by more explicit instruction. MW

Education, T(15-17). *Experiences in Math for Young Children, Second Edition.* Rosalind Charlesworth, Deanna J. Radeloff. Delmar, 1991, vii + 454 pp, $17.67. [ISBN: 0-8273-4630-1] For pre-service and in-service teachers of pre-school and primary age children. Background information on concept development à la Piaget, then sample tasks emphasizing a developmental sequence through naturalistic, informal, and structured learning. MW

Education, T(14-16). *Elementary Mathematics for Teachers, Second Edition.* Donald F. Devine, Judith Olson, Melfried Olson. Wiley, 1991, xvii + 820 pp, $51.95. [ISBN: 0-471-85947-8] Fairly standard treatment with references (though often peripheral) to NCTM *Standards* and recent learning research. Two chapters on problem solving, one on Logo, but little integration of either into the rest of the text. MW

Education, S(16-17). *Math Power in the Community.* Ed: Gerald Kulm. AAAS, 1990, xviii + 217 pp, $9.95 (P), [ISBN: 0-87168-398-9]; *Math Power at Home,* xxiv + 217 pp, $9.95 (P), [ISBN: 0-87168-397-0]; *Math Power in School,* xxii + 235 pp, $9.95 (P); $19.95 set. [ISBN: 0-87168-396-2] Coordinated sets of enrichment activities for middle school students in home, school, and community settings. Leader notes and student sheets, some borrowed from other sources, guide exploration in patterns and symbols; rational numbers; geometric and spatial reasoning; and probability and statistics. Emphasis on problem solving, application, estimation, and conceptual understanding. Permission to copy allows teachers to coordinate efforts by sending home targeted activities. MW

History. *Paul Halmos: Celebrating 50 Years of Mathematics.* Eds: John H. Ewing, F.W. Gehring. Springer-Verlag, 1991, ix + 320 pp, $49. [ISBN: 0-387-97509-8] A *Festschrift* published on the occasion of Halmos's 75th birthday: half biographical reminiscing by Halmos's many friends and colleagues, half mathematical exposition on topics related to Halmos's diverse mathematical interests. Contains many fascinating tales by and about the many men of mathematics who have been influenced by Halmos at different stages of his career. What shines through is not only Halmos's well-known love for mathematics and writing, but also his dedication to teaching and his thorough sense of professionalism. LAS

Foundations, S(16-18), P, L. *Axiomatic Set Theory.* Paul Bernays. Dover, 1991, viii + 227 pp, $6.95 (P). [ISBN: 0-486-66637-9] An unabridged and unaltered republication of the 1968 *Second Edition* of the work first published by North-Holland in 1958. Includes an introduction to the original Zermelo-Fraenkel set theory and a detailed account of the author's independent presentation of a formal system of axiomatic set theory. LCL

Linear Algebra, T(15-16: 1, 2). *Matrix Theory with Applications.* Jack L. Goldberg. McGraw-Hill, 1991, xxi + 520 pp. [ISBN: 0-07-557200-1] Designed to be used as a textbook in an upper-division course on matrix theory for students in engineering and the physical sciences. The first six chapters provide the foundation for a detailed study of the eigenvalue-eigenvector problem; the remaining chapters cover more advanced material on unitary similarities, quadratic forms, and convex polyhedrons. Optional subsections and exercises illustrate the use of MATLAB. AO

Linear Algebra, T(17), S, P, L. *Generalized Inverses of Linear Transformations.* S.L. Campbell, C.D. Meyer, Jr. Dover, 1991, xi + 272 pp, $8.95 (P). [ISBN: 0-486-66693-X] Emphasis placed on utility by presenting many diverse applications, e.g., of the generalized inverse to electrical engineering and of the Drazin inverse to Markov chains. Elementary linear algebra assumed. (1979 Pitman edition, TR, December 1979.) KS

Algebra, T(18: 1, 2), P. *Ring Theory, Student Edition.* Louis H. Rowen. Academic Pr, 1991, xxvii + 623 pp, $55. [ISBN: 0-12-599840-6] An abridged edition of the author's two-volume work (TRs January 1989 and February 1989). More suitable as a graduate-level text in ring theory. Material not needed for proving essential theorems, especially in the later chapters, has been cut. Formating and numbering of results matches the original volumes, enabling the student to find easy access to further reading. LCL

Algebra, P. *Representation Theory of Finite Groups and Finite-Dimensional Algebras.* Eds: G.O. Michler, C.M. Ringel. Progress in Math., V. 95. Birkhäuser Boston, 1991, ix + 520 pp, $86.50. [ISBN: 0-8176-2604-2] Seven survey articles and seventeen research articles presented at the concluding conference in Bielefeld, 1991, for the Deutsche Forschungsgemeinschaft, begun in 1984. JS

Algebra, P. *Computational Aspects of Lie Group Representations and Related Topics.* Ed: A.M. Cohen. CWI Tract, V. 84. Stichting Mathematisch Centrum, 1991, ii + 142 pp, Dfl. 39 (P). [ISBN: 90-6196-395-8] Proceedings of the 1990 Computational Algebra Seminar at CWI, Amsterdam. LC

Calculus, T(13). *Brief Calculus with Applications, Third Edition.* Roland E. Larson, Robert P. Hostetler, Bruce H. Edwards. DC Heath, 1991, xvi + 919 pp, $40 net. [ISBN: 0-669-21767-0] In this edition chapters begin with an overview and end with discussion questions. Calculator exercises and new business applications are added; the treatment of probability is expanded. (*First Edition,* TR, March 1984; *Second Edition,* TR, December 1987.) OJ

Real Analysis, T(15-16: 1, 2), S, L. *Fundamentals of Abstract Analysis.* Andrew M. Gleason. Jones & Bartlett, 1991, xi + 404 pp, $35. [ISBN: 0-86720-209-2] "Were I starting from scratch, I would probably write a different book, but not a very different book." Re-issue of a 1966 Addison-Wesley text (TR, March 1969) that introduces budding mathematics majors to the set-theoretic point of view that has so profoundly shaped modern mathematics. Based

on naive set theory, this "bridge" text moves from sets and logic through construction of real and complex numbers, to metric spaces and analytic functions. A classic text that has withstood well the test of time. LAS

Complex Analysis, S(18), P. *Geometric Function Theory in Several Complex Variables.* Junjiro Noguchi, Takushiro Ochiai. Transl. of Math. Mono., V. 80. AMS, 1990, xi + 282 pp, $78. [ISBN: 0-8218-4533-0] An expanded and rewritten version of the Japanese original published in 1984. A comprehensive introduction, at a high level, to theory, methods, and results of geometric function theory in several complex variables. Chapters treat hyperbolic manifolds, currents and plurisubharmonic functions, meromorphic maps, Nevanlinna theory, and value distribution of holomorphic curves. Presumes familiarity with basics of several complex variables and complex manifolds. PZ

Complex Analysis, T(17-18: 1-3), P**, L.** *Theory of Complex Functions.* Reinhold Remmert. Grad. Texts in Math., V. 122. Transl: Robert B. Burckel. Springer-Verlag, 1991, xix + 453 pp, $59. [ISBN: 0-387-97195-5] A readable, inviting, enthusiastic—sometimes downright exuberant—introduction (though at a high level) to the theory of complex functions of one variable. A veritable treasure trove of mathematical, historical, and cultural gems: e.g., includes brief biographies of six key figures, a photo of Riemann's tomb, line drawings of Euler, Riemann, Cauchy, and Weierstrass, etc., etc. Development traces three-fold approach to subject: Cauchy's, via function theory; Riemann's, via geometric methods; Weierstrass's, via power series. With a goodly number of exercises. A remarkable, generous, and rewarding book. PZ

Differential Equations, P. *Differential Equations: Stability and Control.* Ed: Saber Elaydi. Lect. Notes in Pure & Appl. Math., V. 127. Marcel Dekker, 1991, xvii + 532 pp, $110 (P). [ISBN: 0-8247-8404-9] Proceedings (fifty-two papers) of an international conference held at Colorado College, June 7-10, 1989. In addition to papers on stability and control, the volume contains chapters on recent developments in oscillation theory of differential and difference equations, and the present status of the theory of equations with impulses. LCL

Differential Equations, T(15-16: 1, 2). *Differential Equations with Applications and Historical Notes, Second Edition.* George F. Simmons, John S. Robertson. McGraw-Hill, 1991, xxi + 629 pp, $44.95. [ISBN: 0-07-057540-1] Main changes in this edition include more exercises; new chapters on Fourier series, on partial differential equations, and on numerical methods; new material on operator methods; and engineering applications of the Laplace transform. (*First Edition,* TR, June-July 1972; Extended Reviews, October 1975 and November 1975.) LCL

Differential Equations, T(18: 1, 2), P. *Nonlinear Ordinary Differential Equations and Their Applications.* P.L. Sachdev. Pure & Appl. Math., V. 142. Marcel Dekker, 1991, xviii + 578 pp, $110. [ISBN: 0-8247-8364-6] A high-level, but comparatively leisurely and informal, introduction to theory and applications of nonlinear ordinary differential equations, from the analytic point of view. Chapters 1-4 are elementary; many results are classical. Chapters 5-9 survey modern results and techniques. Applications are made to fluid dynamics, biomathematical problems, and other areas. A wealth of examples and diagrams motivate more theoretical treatments. PZ

Numerical Analysis, T(15-17: 1), C, L*. *Numerical Methods and Software.* David Kahaner, Cleve Moler, Stephen Nash. Prentice Hall, 1989, xii + 495 pp. [ISBN: 0-13-627258-4] Written in a style similar to the now-classic *Computer Methods for Mathematical Computations* by G. Forsythe, *et al* (TR, February 1978). Intermediate in level between a cookbook and a numerical analysis textbook, this volume discusses many of the topics most important for scientists and engineers who will use numerical methods in their work. Topics include computer arithmetic, solution of linear and nonlinear systems of equations, interpolation, numerical quadrature, linear and nonlinear least-squares, data fitting, integration of ordinary differential equations, random number generators and simulation, and the fast Fourier transform. AO

Operator Theory, P. *The Theory of Subnormal Operators.* John B. Conway. Math. Surveys & Mono., V. 36. AMS, 1991, xv + 436 pp, $128. [ISBN: 0-8218-1536-9] A research monograph containing background material and exercises. Recent development of subnormal operator theory "oriented towards rational functions rather than polynomials." Extensive bibliography. MLR

Functional Analysis, S(18), P. *Interpolation Functors and Interpolation Spaces, Volume I.* Yu. A. Brudnyĭ, N. Ya. Krugljak. Math. Lib., V. 47. North-Holland (US Distr: Elsevier Science), 1991, xv + 718 pp, $157. [ISBN: 0-444-88001-1] Classical interpolation theorems, interpolation spaces and interpolation functors, the real method, and selected questions of the theory of the real method. *Volumes II* and *III* to be on the complex method and applications, respectively. Acquaintance with functional analysis and function theory required. KS

Functional Analysis, P. *Orthogonal Schauder Bases.* Taqdir Husain. Pure & Appl. Math., V. 143. Marcel Dekker, 1991, xvii + 283 pp, $99.75. [ISBN: 0-8247-8508-8] "Is each multiplicative linear functional on a complex Fréchet algebra continuous?" This monograph contains an account of recent research on this long-standing open question of E. Michael. MLR

Analysis, P. *Progress in Approximation Theory.* Eds: Paul Nevai, Allan Pinkus. Academic Pr, 1991, xi + 916 pp, $189. [ISBN: 0-12-516750-4] Sixty-two research articles on pure and applied approximation theory and related areas. Note price. LC

Analysis, T(18: 1), S, P, L. *Integral Equations: A Practical Treatment, From Spectral Theory to Applications.* David Porter, David S.G. Stirling. Texts in Appl. Math. Cambridge Univ Pr, 1990, xi + 372 pp, $69.50; $27.95 (P). [ISBN: 0-521-33151-X; 0-521-33742-9] Written for a broader audience interested in applications as well as basic theory; assumes some background in real and functional analysis, differential equations. Offers a readable treatment of such topics as compact and positive operators, approximation methods, and singular integral equations. Plentiful exercises, commentary, appendices, index. JS

Analysis, P. *Inequalities: Fifty Years On from Hardy, Littlewood, and Pólya.* Ed: W. Norrie Everitt. Lect. Notes in Pure & Appl. Math., V. 129. Marcel Dekker, 1991, ix + 283 pp, $99.75 (P). [ISBN: 0-8247-8488-X] Thirteen (of the fourteen) plenary lectures given at an international conference held at the University of Birmingham, England, July 13-17, 1987. The conference was devoted to encouraging the study of inequalities and to acknowledge the indebtedness of the subject to the work of Hardy, Littlewood, and Pólya. LCL

Algebraic Geometry, P. *Number Theory III: Diophantine Geometry.* Ed: Serge Lang. Springer-Verlag, 1991, xiii + 296 pp, $59. [ISBN: 0-387-53004-5] An up-to-date exposition of results and conjectures in the field. The wide-ranging text reaches such topics as heights, Faltings' finiteness theorems, modular curves over Q, Arakelov theory, and diophantine approximations. Proofs are mostly omitted. A valuable reference. SG

Algebraic Geometry, P. *Complex Abelian Varieties and Theta Functions.* George R. Kempf. Universitext. Springer-Verlag, 1991, ix + 100 pp, $29. [ISBN: 0-387-53168-8] An exposition of the analytic approach to complex abelian varieties. The book covers invertible sheaves and sections, and cohomology of abelian varieties, mappings to projective space, theta functions, moduli spaces, modular functions, and Mumford's theta group. SG

Differential Geometry, P. *D-modules and Spherical Representations.* Frédéric V. Bien. Math. Notes, V. 39. Princeton Univ Pr, 1990, 131 pp, $22.50 (P). [ISBN: 0-691-02517-7] The author's Ph.D. thesis (advisor, Joseph Bernstein); the goal is to study the representations of reductive Lie groups which occur in the space of smooth functions on an indefinite symmetric space. LCL

Geometry, S(15-18), P**, L***.** *Unsolved Problems in Geometry.* Hallard T. Croft, Kenneth J. Falconer, Richard K. Guy. Problem Books in Math. Springer-Verlag, 1991, xv + 198 pp, $39.95. [ISBN: 0-387-97506-3] An invaluable resource that will be appreciated at several levels: for research mathematicians there is a supply of problems with bibliography and current status; for general mathematicians, it is a source of interesting reading; and for laymen (and students), it is an introduction to geometry at a research level. The problems are grouped into seven chapters: convexity; polygons, polyhedra, and polytopes; tiling and dissection; packing and covering; combinatorial geometry; finite sets of points; general geometric problems. We are indebted to the authors and other correspondents who have contributed to this volume over the past 30 years; a marvelous book! LCL

Geometry, P. *Vision Geometry.* Eds: Robert A. Melter, Azriel Rosenfeld, Prabir Bhattacharya. Contemp. Math., V. 119. AMS, 1991, xiii + 237 pp, $87 (P). [ISBN: 0-8218-5125-X] Contains fourteen papers from a Special Session on "Geometry Related to Computer Vision" at the AMS meeting in Hoboken, New Jersey, October 20-21, 1989. OJ

Control Theory, T(17-18), P. *Lecture Notes in Control and Information Sciences-153: Control Theory in the Plane.* O. Hájek. Springer-Verlag, 1991, x + 269 pp, $42 (P). [ISBN: 0-387-53553-5] An example-driven introduction to the study of control systems governed by ordinary differential equations. The first three chapters are general in nature; the subsequent four chapters are devoted to topics in the plane, with special reference to the two-dimensional phase plane as state space. Includes many exercises. LCL

Systems Theory, T(16-17: 1), P, L. *Kalman Filtering with Real-Time Applications, Second Edition.* C.K. Chui, G. Chen. Ser. in Inform. Sci., V. 17. Springer-Verlag, 1991, xvi + 195 pp, $39.50 (P). [ISBN: 0-387-18395-7] A concise introduction to the mathematical theory of Kalman filtering, this edition has an expanded section on "real-time system identification," and a brief introduction to wavelet analysis. LCL

Probability, P. *Advances in Probability Distributions with Given Marginals: Beyond the Copulas.* Eds: G. Dall'Aglio, S. Kotz, G. Salinetti. Math. & Its Applic., V. 67. Kluwer Academic, 1991, xviii + 231 pp, $94. [ISBN: 0-7923-1156-5] Contains eleven invited papers given at a conference in Rome, April 1990. A breadth of topics reflects the wide variety of work in the area. TAV

Probability, P. *Probability Theory and its Applications in China.* Eds: Yan Shi-Jian, Yang Chung-Chun, Wang Jia-Gang. Contemp. Math., V. 118. AMS, 1991, xv + 333 pp, $46 (P). [ISBN: 0-8218-5126-8] Nineteen papers, mostly surveys of recent works, collected to demonstrate the breadth of current work in probability in China. The editors hope this collection will help spur collaborative work with Western researchers. TAV

Stochastic Processes, P. *Stochastic Analysis: Liber Amicorum for Moshe Zakai.* Eds: Eddy Mayer-Wolf, Ely Merzbach, Adam Shwartz. Academic Pr, 1991, xx + 532 pp, $69.95. [ISBN:0-12-481005-5] Twenty-nine papers in the fields of stochastic differential equations, nonlinear filtering, two parameter Martingales, Wiener spaces and related topics—all areas of Zakai's interests. Presented in Haifa in January 1991 in honor of Zakai's 65th birthday. TAV

Stochastic Processes, T(16), P*. *The Theory*

of Optimal Stopping. Y.S. Chow, Herbert Robbins, David Siegmund. Dover, 1991, xii + 141 pp, $6.95 (P). [ISBN: 0-486-66650-6] A reprint of the 1971 text (original title: *Great Expectations,* TR, February 1972). The topic remains fresh and fascinating twenty years later. Contains numerous examples, exercises, and a good (if dated) bibliography. TAV

Stochastic Processes, P. *Generalized Diffusion Processes.* N.I. Portenko. Transl. of Math. Mono., V. 83. AMS, 1990, ix + 180 pp, $81. [ISBN: 0-8218-4538-1] The author's (1978) doctoral dissertation, this treatment of diffusion processes in a medium moving in a nonregular way assumes a strong background: Martingale theory, stochastic differential equations, Markov processes, and the main properties of second order partial differential equations of parabolic type. With this background the reader finds a coherent and careful text. Contains a limited bibliography. TAV

Statistics, P. *Nonparametric Functional Estimation and Related Topics.* Ed: George Roussas. NATO ASI Ser. C, V. 335. Kluwer Academic, 1991, xiii + 708 pp, $215. [ISBN: 0-7923-1226-0] Collection of papers presented at a NATO-sponsored Advanced Study Institute on nonparametric functional estimation held in Greece in 1990. Includes sections on curve and functional estimation, parameter selection, regression models, dependent data, time series analysis, and various related topics. List of attendants, index. Note price. JS

Statistics, P, L. *Statistical Reasoning with Imprecise Probabilities.* Peter Walley. Chapman & Hall, 1991, xii + 706 pp, $137. [ISBN: 0-412-28660-2] The author adopts a behavioral interpretation of probability to justify a set of coherence axioms leading to a modification of Bayes theory to allow for imprecise probabilities (including, but not limited to, upper and lower probabilities). Resulting theory is compared and contrasted with Bayes theory. Some discussion on the assessment and elicitation of imprecise probabilities. RWJ

Statistics, T(17), S*, P. *Probability, Statistical Optics, and Data Testing: A Problem Solving Approach, Second Edition.* B. Roy Frieden. Ser. in Inform. Sci., V. 10. Springer-Verlag, 1991, xx + 443 pp, $59 (P). [ISBN: 0-387-53310-9] While the setting is statistical optics, the author's intent is to provide an extensive introduction to the power and applicability of probabilistic and statistical modelling. The stochastic nature of many problems in optics gives a rich and relevant environment for this study. A well-conceived and executed text. TAV

Statistics, T(15-17: 1), L. *Statistical Methods: The Geometric Approach.* David J. Saville, Graham R. Wood. Texts in Stat. Springer-Verlag, 1991, xv + 560 pp, $49.50. [ISBN: 0-387-97517-9] Novel presentation of the main ideas of analysis of variance, analysis of covariance, and regression analysis. Uses a geometric approach, assuming a background of elementary vector geometry and a first course in statistics. Contains numerous examples, which are principally agricultural or biological. Computations are done using Minitab commands. RSK

Computer Literacy, S(14-16), L*. *Computerization and Controversy: Value Conflicts and Social Choices.* Eds: Charles Dunlop, Rob Kling. Academic Pr, 1991, xviii + 758 pp, $34.95 (P). [ISBN: 0-12-224356-0] A thoughtfully selected and carefully edited anthology of nearly fifty readings on a variety of issues pertaining to the role of computers as a powerful factor for social change: technological utopianism, economic and organizational dimensions, transformation of work, social relationships, privacy, security, and ethical responsibilities. Each section opens with a very helpful introduction by the editors complete with references for further reading. LAS

Computer Systems, S, C. *Mathematica Help Stack.* Robert Campbell, *et al.* Variable Symbols (2161 Shattuck Ave., Suite 202, Berkeley, CA 94704-1313; 415-843-8701), 1990, $89. A Macintosh hyper-card application providing a visual tree-structured reference manual for the symbolic computation package *Mathematica,* Version 1.2. Opening "map" groups commands by type (e.g., calculus, number theory, lists); next layer shows options within each type; third layer is a fully active "ref page" for the selected option, giving syntax, description, examples, cross references, and offering an active window for user-added notes. Alternate commands and cross references are instantly available by a mouse-click. Requires 4MB to operate; to simultaneously run *Mathematica* would require much more. LAS

Computer Systems, S. *Mathematica Quick Reference.* Nancy Blachman. Variable Symbols (2161 Shattuck Ave., Suite 202, Berkeley, CA 94704-1313; 415-843-8701), 1990, $14.95. A spiral-bound, slim (4 × 8.5 in.) alphabetic list of all *Mathematica* built-in objects (functions, variables, options) and packages, giving exact syntax and brief description. Handy, easily portable, and very useful. LAS

Computer Systems, S. *UNIX RefGuide.* McNulty Development. Prentice Hall, 1986, iv + 124 pp, (P). [ISBN: 0-13-938957-0] A spiral-bound card file containing a mixture of introductory examples and reference charts for the most common UNIX commands. The thin selection of commands makes it more valuable as a guide (for beginners) than as a reference. Somewhat dated, but still useful. LAS

Artificial Intelligence, P. *Neural Networks: Theory and Applications.* Eds: Richard J. Mammone, Yehoshua Y. Zeevi. Academic Pr, 1991, xv + 355 pp, $39.95. [ISBN: 0-12-467050-4] Papers from a 1990 conference at Rutgers giving overview of features of neural nets and promising approaches. Topics include self organization, learning, vision, character recognition, hybrid nets, nets to solve combinatorial problems. RM

Artificial Intelligence, S(16), L. *Computation & Cognition: Proceedings of the First NEC Research Symposium.* Ed: C.W. Gear. SIAM, 1991, ix + 168

pp, $27.50. [ISBN: 0-89871-272-6] Eight research papers presented at a symposium held in Princeton, New Jersey in May 1989. Under the main theme of models for intelligent computation, it addressed a number of different research efforts in the area of computer design which are aimed at building computers for artificial intelligence and cognitive science. This includes such projects as cellular automata and neural networks. GMS

Computer Science, S(15-16), P, L. *Practical UNIX Security.* Simson Garfinkel, Gene Spafford. O'Reilly & Assoc, 1991, xvii + 481 pp, $29.95 (P). [ISBN:0-937175-72-2] UNIX is one of the most widely used and widely studied operating systems. One of the major complaints against it is that it is a very insecure system that can easily be broken into by the typical "hacker." This book addresses that problem by defining what a secure operating system is and how to achieve relatively high levels within the typical UNIX implementation. GMS

Computer Science, S(16), P, L. *Taking Software Design Seriously: Practical Techniques for Human-Computer Interaction Design.* Ed: John Karat. Academic Pr, 1991, x + 357 pp, $44.50. [ISBN: 0-12-397710-X] Contains seventeen papers presented at a workshop on software engineering in April 1990. All of the papers address the cause of human-computer interface design, especially as it relates to the implementation of practical real world software. Papers also address questions on interface design evaluation, tools, and training techniques. GMS

Computer Science, S(15-16), P, L. *POSIX Programmer's Guide: Writing Portable UNIX Programs with the POSIX.1 Standard.* Donald A. Lewine. O'Reilly & Assoc, 1991, xxvii + 597 pp, $34.95 (P). [ISBN: 0-937175-73-0] POSIX stands for Portable Operating System based on UNIX. It is an attempt to create a single universal UNIX standard so that UNIX-based applications will be portable to a wide range of systems. Book contains a detailed technical description of the POSIX standard as established by OSF, the Open Software Foundation. GMS

Applications, P. *Transactions of the Eighth Army Conference on Applied Mathematics and Computing.* US Army Research Office (POB 12211, Research Triangle Park, NC 27709), 1990, xxi + 925 pp, (P). Technical papers on a wide variety of topics, notably turbulance, variational problems, and symbolic computation, from a conference held at Cornell University during June 1990. LAS

Applications (Biological Science), S(17-18). *Theory of Heart: Biomechanics, Biophysics, and Nonlinear Dynamics of Cardiac Function.* Eds: Leon Glass, Peter Hunter, Andrew McCulloch. Inst. for Nonlinear Sci. Springer-Verlag, 1991, xvii + 611 pp, $59. [ISBN: 0-387-97483-0] A collection of lengthy papers covering diverse topics in cardiology from a physical science perspective. Most of the papers contain little if any mathematics; however, a few demonstrate an impressive command of higher mathematics. MU

Applications (Economics), P. *Control Theory and Dynamic Games in Economic Policy Analysis.* Maria Luisa Petit. Cambridge Univ Pr, 1990, xiv + 338 pp, $49.50. [ISBN: 0-521-38523-7] Mathematical techniques applied to economic policy analysis (stabilization and control of economic systems) based on theory of dynamic deterministic economic models, and continuous time analysis. Addressed to economists. RM

Applications (Physics), T(18: 2), S, P, L. *Lie Algebras, Part 1: Finite and Infinite Dimensional Lie Algebras and Applications in Physics.* G.G.A. Bäuerle, E.A. de Kerf. Stud. in Math. Physics, V. 1. North-Holland (US Distr: Elsevier Science), 1990, xvi + 394 pp, $105.75. [ISBN: 0-444-88776-8] Written to bridge the gap between formal rigorous development of theory of Lie algebras and its availability for application by theoretical physicists; assumes only a background of undergraduate linear algebra (and considerable motivation). Primary goal is the structure theory for complex semi-simple finite-dimensional Lie algebras followed by construction and classification of Kac-Moody algebras. Necessary definitions and sufficient theorems developed to make results intelligible. Few proofs, many references. Numerous remarks, examples, exercises. References, index. JS

Applications (Physics), T(16-18: 1, 2), S, L. *An Introduction to General Relativity.* L.P. Hughston, K.P. Tod. London Math. Soc. Student Texts, V. 5. Cambridge Univ Pr, 1990, 183 pp, $49.50; $19.95 (P). [ISBN: 0-521-32705-9; 0-521-33943-X] It is difficult to imagine a suitable audience for this overly ambitious text. Many significant gaps exist in the development of the mathematical formulae. Thus many readers who are not already fluent with the attendant mathematics will require extensive collateral instruction. Unfortunately, the authors continue the use of the Einstein summation convention, which creates ambiguity and will frustrate many mathematical readers. MU

Applications (Physics), S(17-18). *Mechanics, Analysis and Geometry: 200 Years after Lagrange.* Ed: Mauro Francaviglia. Delta Ser. North-Holland (US Distr: Elsevier Sci), 1991, xi + 559 pp, $137. [ISBN: 0-444-88958-2] Solicited articles by eminent scientists on dynamical systems, integrable systems and quantum groups, analytical mechanics and calculus of variations, global analysis, differential geometry, relativity and field theory, and history of mathematics. MU

Reviewers

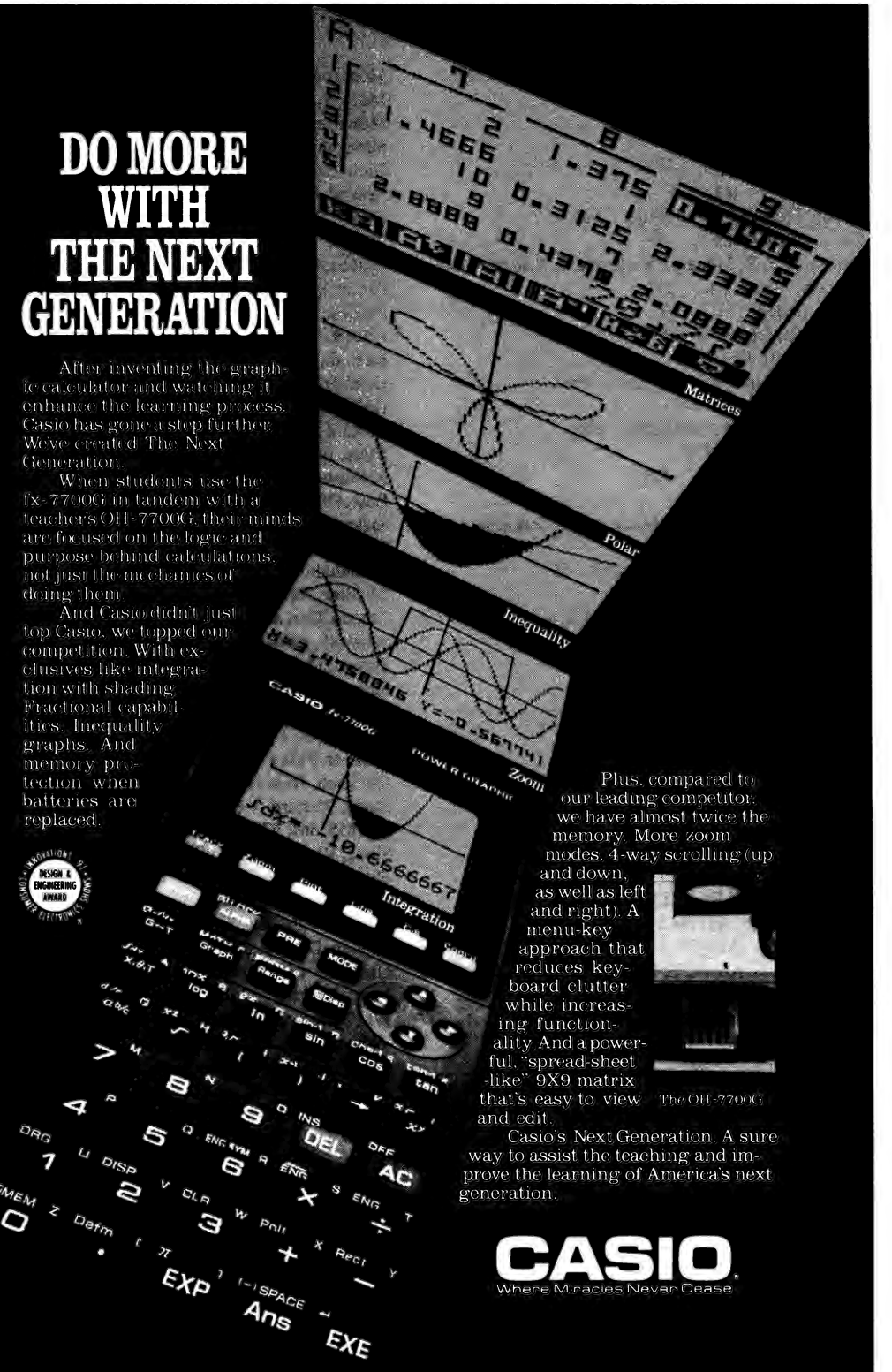
LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; OJ: Ockle Johnson, St. Olaf; RWJ: Roger W. Johnson, Carleton; RSK: Richard S. Kleber, St. Olaf; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; KS: Karen Saxe, Macalester; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

DO MORE WITH THE NEXT GENERATION

After inventing the graphic calculator and watching it enhance the learning process, Casio has gone a step further. We've created The Next Generation.

When students use the fx-7700G in tandem with a teacher's OH-7700G, their minds are focused on the logic and purpose behind calculations, not just the mechanics of doing them.

And Casio didn't just top Casio, we topped our competition. With exclusives like integration with shading. Fractional capabilities. Inequality graphs. And memory protection when batteries are replaced.



Plus, compared to our leading competitor, we have almost twice the memory. More zoom modes. 4-way scrolling (up and down, as well as left and right). A menu-key approach that reduces keyboard clutter while increasing functionality. And a powerful, "spread-sheet-like" 9X9 matrix that's easy to view and edit.

The OH-7700G

Casio's Next Generation. A sure way to assist the teaching and improve the learning of America's next generation.

CASIO

Where Miracles Never Cease

Casio Educational Product Distributors

ABC SCHOOL SUPPLY
800-669-4ABC
(IN GA 404-497-0001)

ALLIED NATIONAL
800-999-8099
(IN MI 313-543-1232)

ARROWHEAD BUSINESS MACHINES
800-234-3396
(IN TX 214-869-0721)
(IN MO 816-861-1113)
(IN CA 213-946-6680)

THE BACH COMPANY
800-248-2224
(IN CA 415-424-0800)

BARDEEN SCHOOL SUPPLY
800-831-0113
(IN NY 315-437-7566)

BECKLEY-CARDY COMPANY
800-446-1477

BHARDS PUBLISHING
800-473-7999
(IN IL 312-642-8657)

CALCULATORS INC.
800-533-9921
(IN MN 800-533-9921)

CAROLINA WHOLESALE
800-521-4600
(IN NC 704-598-8101)

COLBORN SCHOOL SUPPLY INC.
800-548-7031
(IN CO 303-778-1220)
(IN MT 406-245-3158)

COPCO
800-446-7021
(IN OH 800-589-3006)

CUISENAIRE
800-237-3142
(IN NY 914-235-0900)

DALE SEYMOUR PUBLICATIONS
800-872-1100
(IN CA 800-872-1100)

THE DOUGLAS STEWART CO.
800-279-2795
(IN WI 800-279-2795)

E.A.I.
800-272-0272
(IN NJ 201-891-9466)

EDUCATIONAL ELECTRONICS
800-526-9060
(IN MA 617-821-6458)

GLOBAL PRODUCTS
800-633-0633
(IN IL 708-397-4944)

GREYSTONE EDUCATIONAL MATERIALS
800-733-0671
(IN MN 612-430-9857)

KURTZ BROTHERS
800-252-3811
(IN PA 814-765-6561)

LONGINO DISTRIBUTORS
800-633-6224
(IN NC 704-873-3282)

NASCO
800-558-9595
(IN WI 414-563-2446)

NATIONAL AUDIO-VISUAL SUPPLY
800-222-0109
(IN NJ 800-222-0109)

NEMESIS DISTRIBUTING
800-940-7407
(IN FL 305-477-8822)

PENNS VALLEY PUBLISHING/ LEARNING SYSTEMS
800-422-4412
(IN PA 215-855-4948)

SERVCO PACIFIC
(IN HI 808-841-7566)

TAM'S
800-421-5188
(IN CA 800-244-5624)

TAYLOR ELECTRIC
800-558-6970
(IN WI 800-242-8940)

TECHLINE
800-777-3635
(IN VA 703-389-0857)

TROXELL COMMUNICATIONS INC.
800-528-7912
(IN AZ 800-352-7941)

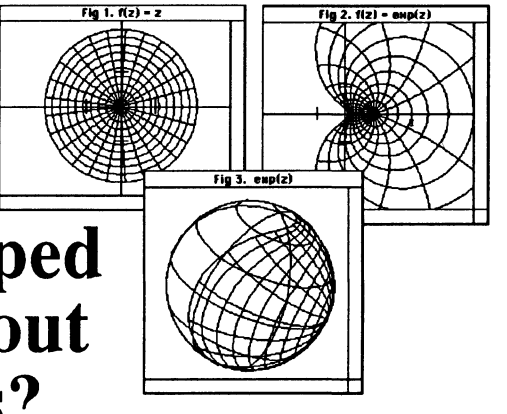
VALLEY BUSINESS MACHINES
(IN UT 801-969-6303)

VISTATECH
800-847-9851
(IN NY 212-254-9851)
(IN CA 213-602-0277)

WHOLESALE ELECTRONIC SUPPLY
800-527-2156
(IN TX 800-441-0145)

CASIO.

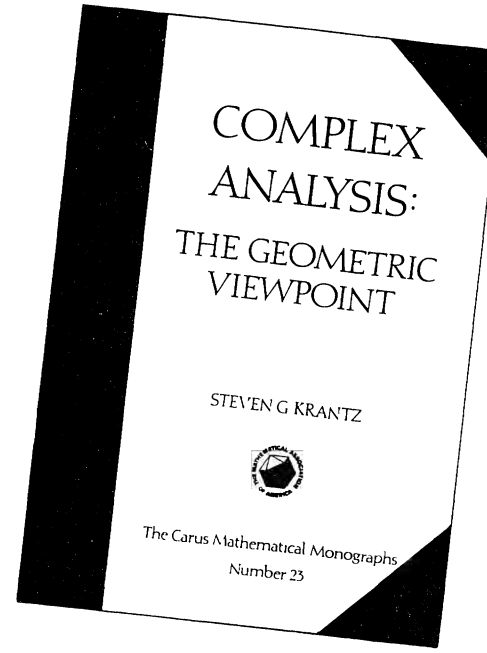
When was the last time a computer program helped you think about mathematics?



Software and video tapes for the student, professional, and anyone who loves mathematics. ***Ask for our latest catalog.***

Lascaux Graphics (602) 544-4229 **(800) 338-0993**
7601 N. Calle Sin Envidia, Suite 31 - Tucson, AZ 85718 USA

New From Steven G. Krantz



COMPLEX ANALYSIS:
THE GEOMETRIC VIEWPOINT

Steven G. Krantz

Geometric methods have been used in complex analysis since the 1930s when Lars Ahlfors discovered that they give a nice way to look at the Schwarz lemma. Since that time they have become a central part of the research activities of complex analysis. However these important techniques have never found their way into a text accessible to a broad audience.

Steven G. Krantz, a leading worker in complex analysis and a well-known mathematical expositor, has written the first book explaining how complex analysis can be studied using methods of geometry. Assuming no background in Riemannian geometry, and only one semester of complex analysis, Krantz explains the role of Hermitian metrics and of curvature in understanding the Schwarz lemma, normal families, Picard's theorems, conformal mappings, and many other topics. A minimum of geometric formalism is used to gain a maximum of geometric and analytic insight. The climax of the book is an introduction to several complex variables from the geometric viewpoint. Poincaré's theorem, that the ball and bidisc are biholomorphically inequivalent, is discussed and proved.

Except for the minimal background requirements, the book is self-contained. A review of relevant topics in the classical theory of one complex variable is provided. The style is light and inviting. The book is a must for anyone with an interest in complex analysis. Take a glance at the main chapter headings and order your copy today

- Principal Ideas of Classical Function Theory,
- Basic Notions of Differential Geometry,
- Curvature and Applications,
- Some New Invariant Metrics,
- A Glimpse of Several Complex Variables

210 pp., 1990, Hardbound,
ISBN 0-88385-026-5

List $22.00 MAA Member $18.50

Catalog Number CAM-23

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

Introducing E.Z. Math, E.Z. Algebra and E.Z. Arithmetic for the HP 48SX

E.Z. Math, E.Z. Algebra and E.Z. Arithmetic are programs for the Hewlett Packard 48SX calculator conceived, written and programmed by Raymond La Barbera and the E.Z. Software Company. Each program comes on a 128K plug-in ROM card accompanied by an easy-to-understand, well-written, detailed manual loaded with lots of specific examples and is designed for use by students, teachers. parents and business people. Each program features an easy-to-use, logically organized, user-friendly interface which enables those who consider themselves to be calculator and computer illiterates, as well as those who don't like to read manuals, to have full access to all program features quickly and easily. Since the HP 48SX is essentially an impressive, sharp-looking, 8 ounce pocket computer, students are easily motivated to take it along with them to study, practice, drill and master math in a study hall, on a train or bus, in a car, on line, on vacation, on a break—in short, for self-study at any time and in any place.

What Can Be Done With E.Z. Math

E.Z. Math effectively solves problems involving graphs, numbers, loans and savings. With E.Z. Math, anyone can:

- Master the entire high school and college graphing curriculum, from algebra to calculus, with 188 families of equations, inequalities, functions, and systems, all arranged in an easy-to-use, user-friendly system of menus to make graphic analysis a snap!
- Get extensive help with calculations involving fractions, whole numbers, complex numbers and number sequences
- Easily do savings and loan calculations and generate complete amortization tables.
- Learn many basic concepts including those involving sets, variables, graphing, solving, numbers, loans and savings

What Can Be Done With E.Z. Algebra

E.Z. Algebra is a comprehensive ninth grade high school basic algebra course as well as a high school and college remedial algebra course that builds a solid algebra foundation. With E.Z. Algebra, anyone can:

- Learn about sets, operations, variables, relations and other concepts essential to a real understanding of algebra.
- Understand the sets of natural numbers, whole numbers, integers, rational numbers and real numbers.
- Master the meaning and properties of the operations of addition, subtraction, multiplication, division, power and root.
- Do all kinds of problems involving algebraic expressions, numerical phrases, equations and inequalities.

What Can Be Done With E.Z. Arithmetic

E.Z. Arithmetic is a comprehensive elementary school basic arithmetic course as well as a high school and college remedial arithmetic course that makes solving most arithmetic problems a snap! With E.Z. Arithmetic, anyone can:

- Learn how to add, subtract, multiply, divide and order whole numbers, fractions, decimals, percents and integers.
- Master the meaning, terminology and conversion methods for whole numbers, fractions, decimals and percents.
- Drill and be graded on endlessly varied, randomly selected sets of problems involving whole numbers, fractions, decimals and integers, with the difficulty level, number of problems, operation and type of number user selectable.

How To Order Copies or Get Further Information

Each E.Z. Software program costs $130.00 ($125.00 retail, plus $5.00 shipping and handling). Take a 10% discount when ordering ten or more units. We accept payment by check, money order, COD, VISA, MC, AE and purchase order. If within 30 days you find that any E.Z. Software program fails to meet your expectations, we'll gladly take back your copy for a prompt, courteous refund. To order copies, either individually or bundled with HP 48SX calculators, please contact:

SMI Corporation, 250 West New Street, Dept MM2, Kingsport, Tennessee 37660
(800) 234-0123 or (615) 378-4821 or (615) 245-8982 (Fax).

ADVENTURES OF A MATHEMATICIAN

S. M. ULAM

WITH A NEW INTRODUCTION BY DANIEL HIRSCH & WILLIAM G. MATHEWS, AND AN EPILOGUE BY FRANÇOISE ULAM & JAN MYCIELSKI

NOW IN PAPERBACK—The autobiography of mathematician Stanislaw Ulam, one of the great minds of the twentieth century, is rich with amazingly prophetic speculations and peppered with lively anecdotes. 384 pages, 25 photographs, $15.95 paper

AFFINE LIE ALGEBRAS, WEIGHT MULTIPLICITIES, AND BRANCHING RULES

VOLUME 1 & VOLUME 2

S. KASS, R. V. MOODY, J. PATERA, & R. SLANSKY

"This is mathematics the way it should be written, with lots of examples. . . . This book is *unique* at this time. . . . The authors have performed a service to the math/physics community."
—Pierre Ramond, University of Florida

Los Alamos Series in Basic and Applied Sciences

840 pages in 2 volumes, 600 tables, $75.00 cloth (2 vol. set)

NEW TITLES IN MATH

At bookstores
or order toll-free
1-800-822-6657.
Visa/MasterCard only.

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY LOS ANGELES
NEW YORK OXFORD

Stories About Maxima and Minima

V.M. Tikhomirov,
Translated by Abe Shenitzer

We are pleased to announce the first volume in a new series of expository books translated from the Russian KVANT Library, published jointly with the American Mathematical Society.

Throughout the history of mathematics, maximum and minimum problems have played an important role. Many beautiful and important problems have appeared in a variety of branches of mathematics and physics. The greatest scientists of the past–Euclid, Archimedes, Heron, the Bernoullis and Newton, took part in seeking solutions to these concrete problems. The solutions stimulated the development of the theory, and, as a result, techniques were elaborated that made possible the solution of a tremendous variety of problems by a single method.

This book presents fifteen "stories" designed to acquaint the reader with the central concepts of the theory of maxima and minima as well as with its illustrious history.

In Part One, the author familiarizes readers with concrete problems that lead to discussion of the work of some of the greatest mathematicians of all time. Part Two introduces a method for solving maximum and minimum problems that originated with Lagrange. While the content of this method has varied constantly, its basic concept has endured for over two centuries. The final story is addressed primarily to those who teach mathematics, for it impinges on the question of how to teach. The author strives to show how the analysis of diverse facts gives rise to a general idea, how this idea is transformed, how it is enriched by new content, and how it remains the same in spite of these changes.

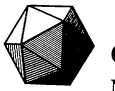
Contents

Ancient maximum and minimum problems: 1. *Why do we solve maximum and minimum problems?*; 2. *The oldest problem–Dido's problem;* 3. *Maxima and minima in nature (optics);* 4. *Maxima and minima in geometry;* 5. *Maxima and minima in algebra and in analysis;* 6. *Kepler's problem;* 7. *The brachistochrone;* 8. *Newton's aerodynamical problem;* **Methods of solution of extremal problems;** 9. *What is a function?;* 10. *What is an extremal problem?;* 11. *Extrema of functions of one variable;* 12. *Extrema of functions of many variables. Lagrange's principle;* 13. *More problem solving;* 14. *What happened later in the theory of extremal problems;* 15. *The fifteenth story, or rather, a discussion.*

187 pp., 1991, Paperbound
ISBN 0-8218-0165-1

List: $23.00 MAA Member: $18.00

Catalog Number: MAXIM



ORDER FROM:
Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036
(FAX) (202) 265-2384

Prepaid orders sent postage & handling free. Visa and Mastercard orders accepted. (Please give the card number and expiration date on credit card orders) We will bill for orders over $10.00.

Princeton

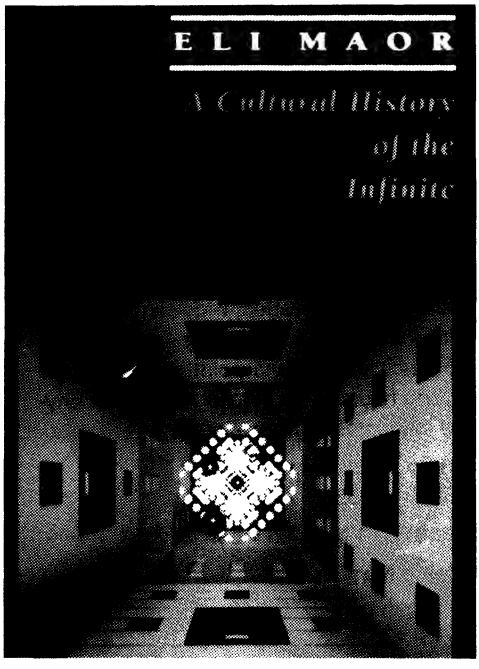
Topics in Noncommutative Geometry

Yuri I. Manin

There is a well-known correspondence between the objects of algebra and geometry: a space gives rise to a function algebra; a vector bundle over the space corresponds to a projective module over this algebra; cohomology can be read off the de Rham complex; and so on. In this book Yuri Manin addresses a variety of instances in which the application of commutative algebra cannot be used to describe geometric objects, emphasizing the recent upsurge of activity in studying noncommutative rings as if they were function rings on "noncommutative spaces."

M. B. Porter Lectures,
Rice University, Department of Mathematics

Cloth: $35.00 ISBN 0-691-08588-9



Introduction to Algebraic and Constructive Quantum Field Theory

John C. Baez, Irving E. Segal, and Zhengfang Zhou

The authors present a rigorous treatment of the first principles of the algebraic and analytic core of quantum field theory. Their aim is to correlate modern mathematical theory with the explanation of the observed process of particle production and of particle-wave duality that heuristic quantum field theory provides. Many topics are treated here in book form for the first time, from the origins of complex structures to the quantization of tachyons and domains of dependence for quantized wave equations.

Princeton Series in Physics
Philip W. Anderson, Arthur S. Wightman, and Sam B. Treiman, Editors

$59.50 ISBN 0-691-08546-3

To Infinity and Beyond

A Cultural History of the Infinite

Eli Maor

Eli Maor describes, with abundant examples, the mathematician's fascination with infinity.

"Fascinating and enjoyable places the ideas of infinity in a cultural context and shows how they have been espoused and molded by mathematics."—*Science*

Over 160 illustrations

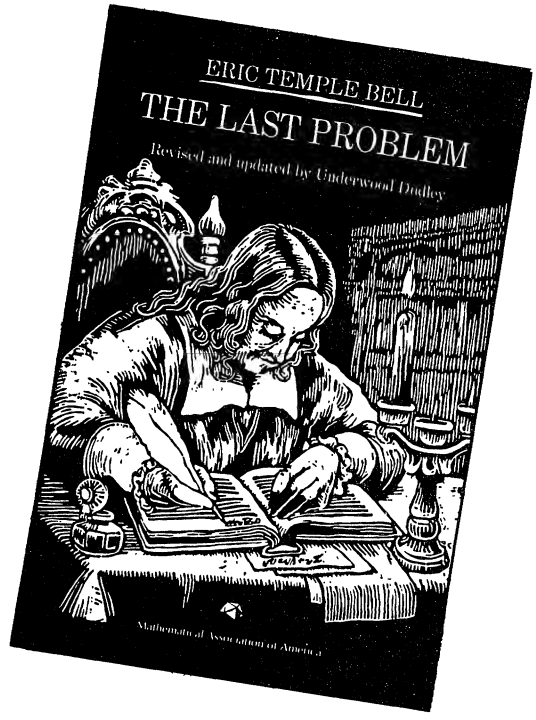
Paper: $16.95 ISBN 0-691-02511-8

Available from Princeton only in the U.S. and Canada

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

Revised and Updated



THE LAST PROBLEM

E. T. Bell
Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990
ISBN-0-88385-451-1

List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

The Ohio State University
invites applications for

TRANSIT

An NSF project to establish **School/University Teams** as regional technology training centers. Regional team training will be provided through summer in-service sessions and academic year follow-up conferences at Ohio State University. Local living expenses with stipend support for pre-college team members is available. Regional teams will help create and/or revise inservice training modules. Regional center teams will begin training teachers as school technology specialists at their regional sites during Summer 1993.

Deadline for completed applications is February 15, 1992.

Write TRANSIT, c/o Frank Demana & Bert Waits, The Ohio State University, Mathematics Department, 231 West 18th Avenue, Columbus, OH 43210.

University of Arizona

GRADUATE FELLOWSHIPS IN THE MATHEMATICAL SCIENCES

Up to 12 fellowships for outstanding new graduate students in the mathematical sciences may be available in 1992-93. Fellowship applicants should be seeking the Ph.D. and planning careers in teaching and/or fundamental research. Anticipated stipends are $12,000 or more for 12 months with both in-state and non-resident tuition waived.

These fellowships are restricted to citizens, permanent residents, or individuals who have established intent to become citizens or permanent residents of the United States. Applications from U.S. women and students belonging to U.S. minority groups are particularly invited. Currently one-fourth of the U.S. graduate students in pure and applied mathematics at Arizona are women.

The University of Arizona has excellent programs in traditional pure and applied mathematics, and is a leading institution in interdisciplinary applied mathematics. This presents a wealth of opportunities for graduate study encompassing such areas as dynamical systems, number theory, computational science, geometry, nonlinear partial differential equations, mathematical physics, probability, and applications of mathematics in the physical, biological, social, and engineering sciences. In addition, outstanding computational facilities for graduate study and research are available to the over 170 graduate students in the mathematical sciences at the University of Arizona.

Fellowship applicants of superior quality will be among the students invited to the **Sixth Annual Workshop for Advanced Undergraduates on Current Ideas in Nonlinear Science, February 29 - March 3, 1992. Limited support is available for attendees.** (The deadline for Workshop applications is February 1, 1992.) The workshop is designed to communicate topics in current active research in three areas: (i) Geometrical and Computational Aspects of Analysis and Number Theory, (ii) Analytical Approaches to Applied Problems, and (iii) Numerical Modelling.

For information and application materials contact:
W. M. Greenlee or T. W. Secomb
Department of Mathematics/Program in Applied Mathematics
University of Arizona - Tucson, AZ 85721 - (602) 621-2068
The University of Arizona is an Equal Opportunity Employer

For College Mathematics Teachers

A SOURCE BOOK FOR COLLEGE MATHEMATICS TEACHING

Alan Schoenfeld, Editor.
Prepared by the Committee on the Undergraduate Teaching of Mathematics

Do you want a broader, deeper, more successful mathematics program? This Source Book points to the resources and perspectives you need.

This book provides the means for improving instruction, and describes the broad spectrum of mathematical skills and perspectives our student should develop. The curriculum recommendations section shows where to look for reports and course resources that will help you in your teaching. Extensive descriptions of advising programs that work is included, along with suggestions for teaching that describe a wide range of instructional techniques. You will learn about how to use computers in your teaching, and how to evaluate your performance as well as that of your students.

Every faculty member concerned about teaching should read this book. Every administrator with responsibility for the quality of mathematics programs should have a copy.

80 pp., 1990, Paper,
ISBN 0-88385-068-0

List $10.00

Catalog Number SRCE

ORDER FROM

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Studies in Numerical Analysis,

Gene H. Golub, Editor
MAA Studies in Mathematics #24

415 pp. Cloth, 1985,
ISBN-0-88385-126-1
List: $46.50
MAA Member: $34.50

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand-held calculator; others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sampling.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Statement of Ownership, Management and Circulation
(Required by 39 U.S.C. 3685)

1A Title of Publication	1B PUBLICATION NO	2 Date of Filing
THE AMERICAN MATHEMATICAL MONTHLY	0 0 0 2 9 8 9 0	Oct. 11, 1991
3 Frequency of Issue	3A No. of Issues Published Annually	3B Annual Subscription Price
Monthly except bi-monthly June/July and Aug/Sept	ten	Library $128 Indiv. Member $32

4 Complete Mailing Address of Known Office of Publication *(Street City, County, State and ZIP+4 Code) (Not printers)*
1529 Eighteenth St., N.W., Washington, D.C. 20036

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*
1529 Eighteenth St., N.W., Washington, D.C. 20036

6 Full Names and Complete Mailing Address of Publisher, Editor and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, D.C. 20036

Editor *(Name and Complete Mailing Address)*
Prof. Herbert S. Wilf, University of Pennsylvania, Dept. of Mathematics, Philadelphia, PA 19104

Managing Editor *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St., N.W., Washington, D.C. 20036

7 Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W., Washington, D.C. 20036

8 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
None	

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 424.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: (Check one)

(1) ☒ Has Not Changed During Preceding 12 Months ☐ Has Changed During Preceding 12 Months (2)

If changed, publisher must submit explanation of change with this statement.

10 Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No. Copies *(Net Press Run)*	23,893	22,477
B Paid and/or Requested Circulation 1 Sales through dealers and carriers, street vendors and counter sales	0	0
2 Mail Subscription *(Paid and/or requested)*	19,431	18,619
C Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	19,431	18,619
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	888	1,127
E Total Distribution *(Sum of C and D)*	20,277	19,746
F Copies Not Distributed 1 Office use, left over, unaccounted, spoiled after printing	3,616	2,731
2 Return from News Agents	0	0
G TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	23,893	22,477

11 I certify that the statements made by me above are correct and complete

Signature and Title of Editor, Publisher, Business Manager, or Owner
Kay Aymont, Membership Manager

PS Form 3526, January 1991 *(See instructions on reverse)*

DERIVE® a compact card.



DERIVE® , A Mathematical Assistant is now available for palmtops through 486-based PCs.

The DERIVE® program solves both symbolic and numeric problems, and it plots beautifully too.

- Symbolic math from algebra through calculus.
- Plots in both 2-D and 3-D.
- Simple, letter-driven menu interface.
- Solves equations exactly.
- Understands vectors and matrices.
- Split or overlay algebra and plot windows.
- Displays accepted math notation.
- Performs arithmetic to thousands of digits.
- Simplifies, factors and expands expressions.
- Does exponential, logarithmic, trigonometric, hyperbolic and probability functions.
- Taylor and Fourier series approximations.
- Permits recursive and iterative programming.
- Can generate Fortran, Pascal and Basic statements.

System requirements

PC version: MS-DOS 2.1 or later, only 512Kb RAM and one 3.5" or 5.25" disk drive. Suggested retail price is $250.

ROM-card version: Hewlett-Packard 95LX Palmtop computer. Suggested retail price is $289.

Contact Soft Warehouse for a list of dealers. Or, ask at your local computer store, software store or HP calculator dealer. Dealer inquires are welcome.

DERIVE is a registered trademark of Soft Warehouse, Inc.

2000 Years of Mathematical Knowledge on a Disk

Soft Warehouse INC
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waialae Avenue
Suite 304 • Honolulu, HI, USA 96816-3236
Phone: (808) 734-5801 • Fax: (808) 735-1105

MATHEMATICS BOOKS FROM BCS

NUMERICAL LINEAR ALGEBRA by Willy Brandal (1991)

Designed as an undergraduate text. Carefully written with the appropriate level of rigor to make it work. Various algorithms are mathematically developed and compared. Care is taken with regard to which results hold in the field of reals and which hold in the complex numbers. Topics include: Gaussian elimination, LU decomposition, Cholesky decomposition, the Gauss-Seidel, power, and QR methods, Aitken acceleration, and others. Includes over 350 exercises distributed at the end of sections, and 5 computer assignments.

ISBN 0-914351-05-2 Paperback only 212pp US$30.00

Note. In order to keep the cost to the student as reasonable as possible:

1) We do not intend to periodically float new editions.

2) We do not send examination copies— we think your library (or department) should purchase a copy.

Translations into English of classics, both old and new.

FUNDAMENTALS OF GRAPH THEORY by A.A.Zykov (1990)

Translation, assisted by the author, of the 1987 Russian edition (M.:Nauka.Gl.red.Fiz.-mat.lit.). Includes a Glossary-Index-Bibliography that should be of immense value to the researcher and student of Graph Theory. An impressive number of exercises of varying difficulty are included at the end of each section. This was written as a textbook, and is admirable suited for that purpose. A sample of some topics included in this vast work: Simple, labeled, multi-, topological and directed graphs; numerous invariants— optimum and critical graphs; the isomorphism problem and Vizing's construction; reconstruction problems; matchings; embeddings in surfaces, planarity; Hadwiger's conjecture; various colorings; perfect graphs; reachability; kernels; and a great many others. There is an Appendix on Boolean methods in graph theory.

ISBN 0-914351-04-4 Paperback only 371pp US$48.00

THEORY OF CONVEX BODIES by T.Bonnesen and W.Fenchel (1987)

Originally published as volume 3 of Springer-Verlag's "Ergebnisse" series in 1934.

ISBN 0-914351-02-8 Paperback only 183pp US$30.00

GRAPH THEORY:A Development from the 4-Color Problem by Martin Aigner (1987)

An exceptionally readable and aptly titled development of graph theory. The original German edition was published in 1984 by B.G.Teubner.

ISBN 0-914351-03-6 Paperback only 233pp US$36.00

TOPOLOGY OF SURFACES by André Gramain (1984)

A beautiful introduction to the concepts and methods of differential topology. Originally published in 1974 by Presses Univeritaires de France.

ISBN 0-914351-01-X Paperback only 227pp US$24.00

KNOT THEORY by Kurt Reidemeister (1983)

A translation of Springer-Verlag's 1932 German edition.

ISBN 0-914351-00-1 Paperback only 158pp US$20.00

Individual members of the American Mathematics Society or Association who cite this ad may take a discount of 10% from the above prices.

Sorry, but we can only accept prepaid orders from individuals and do not accept credit cards. Orders from Canada, Mexico and the USA will be sent 4th class bookrate. Please include $1.75 for postage and handling for the first book, and $.50 for each additional book up to a maximum postage and handling fee of $4.75. Foreign orders must be paid in US dollars.

If you insist, we will ship UPS with the following shipping and handling fees: $5.00 for the first book, and $.75 for each additional book up to a maximum shipping and handling fee of $9.50.

Foreign orders except from Canada or Mexico please add the following postage and handling charges: $3.30 (Surface Mail) for the first book and $1.00 for each additional book. Foreign Airmail will be at cost plus $1.00 per order (cost is approximately $12.00 for the first book and $6.00 for each additional book).

BCS ASSOCIATES

P.O.Box 3614 University Station
Moscow, ID 83843–0478